Studia Scientiarum Mathematicarum Hungarica

VOLUME 28
NUMBERS 1–2
1993

AKADÉMIAI KIADÓ, BUDAPEST

STUDIA SCIENTIARUM MATHEMATICARUM HUNGARICA

A QUARTERLY OF THE HUNGARIAN ACADEMY OF SCIENCES

Studia Scientiarum Mathematicarum Hungarica publishes original papers on mathematics mainly in English, but also in German, French and Russian. It is published in yearly volumes of four issues (mostly double numbers published semiannually) by

AKADÉMIAI KIADÓ
Publishing House of the Hungarian Academy of Sciences
H–1117 Budapest, Prielle Kornélia u. 19–35

Manuscripts and editorial correspondence should be addressed to

J. Merza
Managing Editor

P.O. Box 127
H–1364 Budapest

Tel.: (36)(1) 118–2875 Fax: (36)(1) 117–7166
e-mail: h3299mer @ ella.hu

Subscription information

Subscription price for Volume 28 (1993) in 4 issues: $ 88.00 including normal postage, airmail delivery plus $ 20.00

Orders should be addressed to

AKADÉMIAI KIADÓ
P.O.Box 245
H–1519 Budapest

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 1–15

GENERALIZED P.P. RINGS AND RINGS OF π-REGULAR QUOTIENTS

R. GONCHIGDORZH

The recent papers [8], [10] and [6] have been devoted to the study of noncommutative and commutative generalized p.p. rings with identity. In the noncommutative case the existence of classical right rings of quotients was assumed. In the present paper we continue these investigations for noncommutative normal generalized p.p. rings not necessarily having an identity element. In this case we cannot assume existence of the classical ring of quotients, because it may happen that such a ring has no cancellative element. So, it is a natural question to find a construction of ring-extensions for generalized p.p. normal rings not necessarily having identity which can be considered as a substitution of the construction of the classical rings of quotients. We shall present a construction of a ring of right π-regular quotients of a normal generalized p.p. ring. This is a generalization of the rings of right regular quotients of reduced rings introduced by the author in [4]. The latter one was used in [5] for characterizing semihereditary and hereditary reduced rings without assuming the existence of an identity element (similar results for semihereditary and hereditary rings with identity were obtained by M. Ohori in [10]).

The first section is preliminaries and there we recall some definitions and results needed later.

In Section 2 we shall define rings of right π-regular quotients of normal generalized p.p. rings and we shall give some sufficient and necessary conditions for the existence of rings of right π-regular quotients and corollaries for rings with identity. These results can be considered as characterizations of normal generalized p.p. rings with rings of right π-regular quotients. So, there is some generalization of results in [8].

In the third section we shall consider normal generalized p.p. rings with Köthe radical and there a characterization will be obtained for a normal generalized p.p. ring having a ring of right π-regular quotients with Pierce stalks which are local rings.

The last section is devoted to a characterization of normal generalized p.p. rings with π-regular rings of right π-regular quotients. This is a generalization of some results in [10].

1980 *Mathematics Subject Classification* (1985 Revision). Primary 16A08; Secondary 16A30.

Key words and phrases. Generalized p.p. ring, normal ring, annihilator, Pierce sheave, the ring of π-regular quotients, Ore ring.

Akadémiai Kiadó, Budapest

1. Preliminaries

Throughout this paper, all rings considered are associative and without assuming the existence of identity elements, unless otherwise indicated. Let R be a ring and S be a subset of R. Then $r_R(S)$ and $\ell_R(S)$ denote the right and left annihilator of S in R, respectively. $\operatorname{ann}_R(S)$ will stand for $r_R(S)$ when $r_R(S) = \ell_R(S)$. $E(R)$ and $B(R)$ denote the set of all idempotents and the set of all central idempotents of the ring R, respectively. A ring R is said to be a normal ring if $E(R) = B(R)$, i.e. if every idempotent of R is central.

DEFINITION 1.1. A ring R is said to be a *right generalized p.p. ring* if $E(R)R = R$ and for any $a \in R$ there exist a positive integer n and an idempotent $e \in E(R)$ with $r_R(a^n) = r_R(e)$. A *left generalized p.p. ring* is defined analogously. A ring R is a *generalized p.p. ring* (abbreviated g.p.p. ring) if it is both right and left generalized p.p. ring.

It is clear that if the ring R has an identity this definition coincides with the usual one given in [8], [10]. An extension of the concept of g.p.p. ring to rings without identity was made in [7]. There it was not assumed that $E(R)R = R$ and of course, in such a ring some nil direct summand may occur. Nil rings, however, can be considered as a trivial case of g.p.p. rings. Therefore, in our definition of a g.p.p. ring R we have supposed the condition $E(R)R = R$ for excluding such trivial cases.

LEMMA 1.2 (cf. [8, Corollary 4]). *Let R be a normal ring. Then R is g.p.p. ring if and only if $R \cdot E(R) = R$ and for any $a \in R$ there exist a positive integer n and an idempotent $e \in E(R)$ such that for every $K \geqq n$ $\operatorname{ann}_R(a^K) = \operatorname{ann}_R(e)$ and $a^K e = a^K$.*

PROOF. The necessity is obvious.

Let R be a g.p.p. ring and a be an arbitrary element of R. Then there are integers $m, n > 0$ and idempotents $e, f \in E(R)$ such that $r_R(a_n) = r_R(e)$, $\ell_R(a^m) = r_R(f)$. Then for all $K \geqq \max\{n, m\}$ we have $r_R(a^K) = \operatorname{ann}_R(e)$ and $\ell_R(a^K) = \operatorname{ann}_R(f)$ (cf. [8, Lemma 3]). For, let $a^{n+1}b = 0$, $b \in R$. Then $0 = eab = aeb$ and hence $a^n eb = 0$, $r_R(a^{n+1}) = r_R(a^n)$. In a similar way we have $\ell_R(a^{m+1}) = \ell_R(a^m)$.

Now, let $r_R(a^K) = \operatorname{ann}_R(e)$, $\ell_R(a^K) = \operatorname{ann}_R(f)$. Then $0 = a^K(f - ef) = (f - ef)a^K$, and hence $f(f - ef) = 0$, $f = ef$. Similarly we have $e = ef$ and therefore $e = f$ and $r_R(a^K) = \operatorname{ann}_R(e) = \ell_R(a^K)$, i.e. $\operatorname{ann}_R(a^K) = \operatorname{ann}_R(e)$. Because $E(R) \cdot R = R$, we have an idempotent $e' \in E(R)$ such that $a^K - ea^K = e'(a^K - ea^K) = (e' - e'e)a^K$. Since $e(e' - ee') = 0$, we have $0 = (e' - ee')a^K = a^K - ea^K$ and so, $a^K = ea^K$.

DEFINITION 1.3. An element $r \in R$ is called *reduced* if there is a central idempotent $e \in B(R)$ such that $\operatorname{ann}_R(r) = \operatorname{ann}_R(e)$ and $r = er$. In this case it is easy to verify that the central idempotent e is unique and we call it *associated idempotent* of the reduced element r.

COROLLARY 1.4 (cf. [9, Theorem 1]). *Let R be a normal ring. Then the following are equivalent:*

(1) *R is g.p.p. ring;*

(2) a) $E(R) \cdot R = R$,

b) *every $a \in R$ can be written in the form $a = r + n$ where r is a reduced element and n is a nilpotent element, and the sum is orthogonal, i.e. $rRn = nRr = (0)$.*

PROOF. (1) $\Rightarrow$ (2): Let R be a g.p.p. ring. Then by Lemma 1.2 for some integer $K > 0$ a^K is reduced element with an associated idempotent $e \in B(R)$. Putting $r = ea$ and $n = a - ea$ we have $rRn = nRr = 0$ and $(ea)^K = ea^K = a^K = (ea + a - ea)^K = (ea)^K + n^K$. Hence n is nilpotent and, of course, r is reduced.

The implication (2) $\Rightarrow$ (1) is obvious, because if r is a reduced element, then for every integer $K > 0$ r^K is reduced.

In the rest of this section we shall give some notations, general remarks for sheaves and Pierce sheaves of rings. More details one can find in [1], [3] or in [11], [12].

Let X be a topological space with the set $T(X)$ of open subsets. Then $T(X)$ is a category with morphisms being inclusions of open subsets of X. Every contravariant functor $P: T(X) \to \mathrm{SET}$, where SET is the category of sets, is called *presheaf* on X. So, if a presheaf P on X is given, for each open subset $U \subseteq X$ we have a set $P(U)$ and for every pair of open subsets $V, U \subseteq X$ with $V \subseteq U$ there is a *restriction map* $\varrho_V^U: P(U) \to P(V)$ and these restriction maps satisfy the natural conditions of compatibility.

Presheaf P on X is called a *sheaf* on X if for an arbitrary $U \in T(X)$, for given an open cover $\{U_i \mid i \in I\}$ of U and a family of elements $s_i \in P(U_i)$, $i \in I$, such that for each pair (i, j) we have $\varrho^{U_i}_{U_i \cap U_j}(s_i) = \varrho^{U_j}_{U_i \cap U_j}(s_j)$, there exists a unique element $s \in P(U)$ with $\varrho^U_{U_i}(s_i) = s$ for all $i \in I$. The *stalk* P_x of a presheaf P at a point x of X is defined as the colimit $P_x = \varinjlim_{U \ni x} P(U)$ of the sets $P(U)$ as U ranges over all open neighbourhoods of x which form a filter in $T(X)$. If $s \in P(U)$ for some neighbourhood $U \in T(X)$ of $x \in X$, we write s_x for the image of s in P_x and call it the *germ* of s at x. The collection of the stalks $\{P_x \mid x \in X\}$ is an X-indexed family of sets and we write $S(P)$ for the disjoint union of the P_x with its *canonical projections* $\pi: S(P) \to X$. We define a topology on $S(P)$ declare $V \subseteq S(P)$ to be open if for all $U \in T(X)$ and all $s \in P(u)$, the set $\{x \in U \mid s_x \in V\}$ is open in X. We write $\Gamma(U, S(P))$ for the set of all continuous *partial sections* $\varphi: U \to S(P)$ of the projection $\pi: S(P) \to X$ (i.e. $\varphi \circ \pi$ is inclusion of U in X). A *global section* is that a section with $U = X$. Then if $s \in P(U)$ for some U, the map $x \to s_x$ defines a partial continuous section $\hat{s}: U \to S(P)$ over U.

Now, let R be a ring and $B(R)$ be the set of all central idempotents of R.

Then $B(R)$ forms a Boolean ring under the following operations

$$e \oplus f = e + f - 2ef$$
$$e \cdot f = ef.$$

The Boolean spectrum spec $B(R)$ of the ring R consisting of all prime ideals of $B(R)$ with the Zarisky topology is a totally disconnected, local compact Hausdorff space (i.e. Stone space) and we denote it by $X(R)$. An important property of the space $X(R)$, the *partition property* of $X(R)$ can be formulated as follows.

LEMMA 1.5. *Let U be an open compact subset of $X(R)$ and $\{U_\alpha, \alpha \in I\}$ be an open cover of U. Then there exists a finite set $\{V_i, i = 1, 2, \dots, n\}$ of open compact subsets in $X(R)$ such that*

i) *for every $i \leqq n$ there is an index $\alpha \in I$ with $V_i \subseteq U_\alpha$,*

ii) $V_i \cap V_j = \emptyset$ *if* $i \neq j$,

iii) $U = \bigcup_{i=1}^{n} V_i$.

We define a presheaf on $X(R)$ as follows. Let U be an open subset in $X(R)$. Then we write $P(U) = R/\bigcap_{x \in U} xR$ (we note that x is a prime ideal in $B(R)$ and xR is an ideal in R) and we define

$$\varrho_V^U : R/\bigcap_{x \in U} xR \to R/\bigcap_{y \in V} yR, \quad \text{where} \quad V \subseteq U,$$

to be the canonical projection. It is easy to get that, in fact, this presheaf is a sheaf and a calculation gives that the stalks $P_x = R/xR$ for all $x \in X(R)$. We shall write R_x instead of P_x. This sheaf P is called a *Pierce sheaf* of the ring R and the stalks are called the *Pierce stalks* of R. An adaptation of [11, Proposition 1.1] gives

LEMMA 1.6. *Let R be a ring with $B(R) \cdot R = R$. Then every Pierce stalk R_x of R is a ring with identity. Moreover, each Pierce stalk R_x is indecomposable (i.e. without nontrivial central idempotents) if and only if R is a normal ring.*

In the rest, we suppose that R is a normal ring with $B(R)R = R$. Let $S(R) = \bigcup_{x \in X(R)} R_x$ be the display space and $\hat{r} : X(R) \to S(R)$ be a map defined by an element $r \in R$ by $\hat{r}(x) = r_x$ for all $x \in X(R)$. Then $\hat{r}$ is a continuous global section of the Pierce sheaf of R and the Pierce sheaf $P(R) = (S(R), X(R))$ is a *sheaf of rings R_x, $x \in X(R)$ with identity* in the sense of [3, Definition 2.2]. Moreover, the set of all global sections $\Gamma = \Gamma(X(R), S(R))$ forms a ring under naturally defined ring operations and a map $i : R \to \Gamma$ with $i(r) = \hat{r}$ is a monomorphism of rings. Hence we can identify the ring R with the subring $i(R) = \hat{R} = \{\hat{r} \mid r \in R\}$. It is well-known that $\hat{R} = \Gamma$ if and only

if R has an identity element. Let σ be a global (continuous) section in Γ. Then $\operatorname{supp}\sigma$ denotes the *support* of σ, i.e. $\operatorname{supp}\sigma = \{x \in X(R) \mid \sigma(x) \neq 0_x\}$. We shall end this section with a useful remark: a subset $U \subset X(R)$ is open and compact if and only if $U = \operatorname{supp}\hat{e}$ for an idempotent $e \in B(R)$.

2. The ring of π-regular quotients of normal g.p.p. rings

In order to define the ring of π-regular quotients of normal g.p.p. rings, at first we shall establish some facts about Pierce stalks of normal g.p.p. rings, about reduced elements and about some extensions of normal g.p.p. rings.

LEMMA 2.1. *Let R be a normal g.p.p. ring. Then for each $x \in X(R)$ every zero divisor of the Pierce stalk R_x is nilpotent.*

PROOF. Let $r_x \in R_x$. By Lemma 1.2 for a positive integer m and an idempotent $e \in B(R)$ we have $r^m e = r^m$, $\operatorname{ann}_R(r^m) = \operatorname{ann}_R(e)$. So if $e_x = 0_x$, then $r_x^m = r_x^m e_x = 0_x$ and r_x is nilpotent. Let $e_x \neq 0_x$ and $r_x \cdot s_x = 0_x$ for an element $s_x \in R_x$. Then $rsf = 0$ for some idempotent $f \in B(R)$ with $f_x \neq 0_x$. Moreover, also $ef = f$ holds. Hence $0 = (sf)e = sf$ and, in particular, $0_x = s_x f_x = s_x$. Therefore, r_x is right non-zero divisor in R_x. Similarly we can show that r_x is a left non-zero divisor in R_x.

COROLLARY 2.2. *Let R be a normal g.p.p. ring. Then an element $r \in R$ is reduced if and only if for every $x \in \operatorname{supp} r$ r_x is a non-zero divisor in R_x and in this case* $\operatorname{supp} r$ *is open compact in $X(R)$. Moreover,* $\operatorname{ann}_R(r) = \operatorname{ann}_R(e)$ *if and only if* $\operatorname{supp} r = \operatorname{supp} e$.

PROOF. Obvious.

DEFINITION 2.3. Let R and $\tilde{R}$ be rings with $R \subset \tilde{R}$. Then $\tilde{R}$ is called a *unital extension* of R, if for every element $\tilde{r} \in \tilde{R}$ $E(R)\tilde{r} \neq 0$, $\tilde{r}E(R) \neq 0$ whenever $\tilde{r} \neq 0$.

LEMMA 2.4. *Let R be a normal g.p.p. ring with a unital extension $\tilde{R}$. Then for every element r of R there exists at most one element $r^* \in \tilde{R}$ such that*

$$(*) \qquad r^* r^{m+1} = r^m, \quad \operatorname{ann}_{\tilde{R}}(r^*) = \operatorname{ann}_{\tilde{R}}(r^m) = \operatorname{ann}_{\tilde{R}}(e)$$

for some integer $m > 0$ and idempotent $e \in B(R)$.

If there exists such an element $r^* \in R$ then we call it the *π-regular inverse of the element r* and throughout this paper the notation r^* will be used only in this sense.

PROOF. Let r^*, $r^\# \in R$ satisfy Condition $(*)$ for some element $r \in R$. In view of Lemma 1.2 we may suppose that the integer and the idempotent

appearing in $(*)$ are the same. So we have $(r^* - r^\#)r^{m+1} = 0$ and hence $(r^* - r^\#)e = 0$ (Lemma 1.2). Moreover, if f is an arbitrary idempotent of R then $(r^* - r^\#)(f - ef) = 0$, since $e(f - ef) = 0$. Therefore, we have

$$(r^* - r^\#)f = (r^* - r^\#)(f - ef + ef) = 0$$

and hence $(r^* - r^\#)B(R) = 0$. Because $\bar{R}$ is a unital extension of R the last equality yields $r^* = r^\#$.

It is clear that in any unital extension every nilpotent element has a π-regular inverse which is zero. We know that every element of a normal g.p.p. ring is an orthogonal sum of a reduced element and a nilpotent element (Corollary 1.4). Moreover, Lemma 2.1 and Corollary 2.2 imply that these reduced and nilpotent parts are uniquely determined by the element. So we have

LEMMA 2.5. *Let R be a normal g.p.p. ring and r be an element of R with orthogonal decomposition $r = a + b$, where a is a reduced, b is a nilpotent element. Then the element r has a π-regular inverse r^* in unital extension if and only if the element a has a^* in this extension and $a^* = r^*$.*

DEFINITION 2.6. Let R be a normal g.p.p. ring and Q be a unital extension of R. Then Q is said to be the *ring of right π-regular quotients* of R if the following conditions are satisfied:

(i) every element $r \in R$ has π-regular inverse r^* in Q,

(ii) every element $q \in Q$ has a form $q = rs^*$ for some elements $r, s \in R$.

The ring of right π-regular quotients Q of the ring R (for brevity we denote it by $Q^r_\pi(R)$) has the following universal property, if it exists: Let P be a unital extension of R such that every element of R has π-regular inverse in P. Then there exists a unique homomorphism

$$\varphi\colon Q^r_\pi(R) \to P$$

such that $\varphi(r) = r$ for all $r \in R$. Therefore, $Q^r_\pi(R)$ is unique up to isomorphism. Moreover, we can analogously define the ring of left π-regular quotients $Q^\ell_\pi(R)$ of R. The above universal property says that if both $Q^r_\pi(R)$ and $Q^\ell_\pi(R)$ exist, then they are naturally isomorphic.

Before formulating the main result of this section in view of Lemma 2.5 we notice that Conditions (i) and (ii) in Definition 2.6 can be changed by the following conditions:

(i)′ every reduced element of R has a π-regular inverse in Q,

(ii)′ every element $q \in Q$ has the form $q = rs^*$ for some $r, s \in R$ where s is reduced.

THEOREM 2.7. *Let R be a normal ring. Then the following are equivalent:*

(1) *R is a g.p.p. ring having a ring of right π-regular quotients,*

(2) a) *every Pierce stalk R_x of R is a right Ore ring with identity and each zero divisor in R_x is nilpotent,*

b) *for every element $r \in R$ the set*

$$\{x \in X(R),\ r_x \text{ is non-zero divisor}\}$$

is open compact in $X(R)$ and $\operatorname{supp} r \subseteqq \operatorname{supp} f$ *for some $f \in B(R)$.*

(3) *R is g.p.p. ring with right π-Ore condition: that is for every pair of elements $a, b \in R$ where b is reduced, there exist $c, d \in R$ such that*

$$ad = bc, \quad \operatorname{ann}_R(d) = \operatorname{ann}_R(b).$$

PROOF. (1) $\Rightarrow$ (2): a) By Lemma 2.1 it is sufficient to prove that R_x is right Ore ring for all $x \in X(R)$. For, let $a_x, b_x \in R_x$ be non-zero elements and let b_x be a non-zero divisor in R_x. Then we have $b^* a = cd^*$ for some $c, d \in \in R$ and by Lemma 2.5 we conclude that both b and d are reduced. Hence $bb^* ad = bcd^* d$ and noting that $b_x b_x^* = 1_x$, $d_x^* d_x = 1_x$ (Corollary 2.2) we have the right Ore condition: $a_x d_x = b_x c_x$, where d_x is non-zero divisor in R_x.

Condition (2), b) follows immediately from Definition 1.1 and Corollaries 1.4 and 2.2.

(2) $\Rightarrow$ (3). At first we prove that R is g.p.p. ring. Let r be an arbitrary element of R. By the assumption there are idempotents e, f such that

$$\operatorname{supp} e = \{x \in X(R),\ r_x \text{ is non-zero divisor}\}$$

and $\operatorname{supp} r \subseteqq \operatorname{supp} f$. Since $\operatorname{ann}_R(re) = \operatorname{ann}_R(e)$ and $rf = r$, $r = re + r(f - e)$, we see that re is a reduced element and $reRr(f - e) = 0$. Now we prove that $r' = r(f - e)$ is nilpotent. Since $\operatorname{supp}(f - e)$ is open compact and r'_x is nilpotent for every $x \in \operatorname{supp}(f - e)$ by the standard Pierce sheave argument (cf. Lemma 1.5) we get an orthogonal set of idempotents $\{e_1, \ldots, e_n\}$ with $e_1 + e_2 + \ldots + e_n = f - e$ and integers K_i, $i = 1, 2, \ldots, n$ such that $(r' e_i)^{K_i} = 0$ for all i. Then taking $K = \max\{K_i \mid i = 1, 2, \ldots, n\}$ we have $(r')^K = 0$, as desired.

Now we will test the π-Ore condition. We take elements $a, b \in R$, where b is reduced. By Corollary 2.2 for every $x \in \operatorname{supp} b$ b_x is non-zero divisor in R_x, and hence by our assumption there are elements $c_x, d_x \in R_x$, where d_x is a non-zero divisor in R_x, such that $a_x d_x = b_x c_x$. Noting that if r_x is non-zero divisor in R_x we can assume that r is reduced (see Corollary 1.4) and using again the standard Pierce sheave argument we get an orthogonal set $\{e_1, \ldots, e_n\}$ of idempotents and a set $\{c_1, \ldots, c_n, d_1, \ldots, d_n\}$ of elements of R such that each $d_i e_i$ is reduced and $ad_i e_i = bc_i e_i$, $e_1 + e_2 + \ldots + e_n = e$, where $\operatorname{supp} e = \operatorname{supp} b$. Hence, putting $d = \sum_{i=1}^{n} d_i e_i$, $c = \sum_{i=1}^{n} c_i e_i$ we get the desired equality $ad = bc$ with $\operatorname{ann}_R(d) = \operatorname{ann}_R(b)$, where d is reduced (see Corollary 2.2).

The implication (3) $\Rightarrow$ (2) is obvious (see the proof of (1) $\Rightarrow$ (2)).

Now we prove that Condition (1) follows from Conditions (2) and (3). Assume Conditions (2) and (3). It is well-known that R is a subdirect product of its Pierce stalks R_x, $x \in X(R)$. So, for every reduced element $r \in R$ we can construct a unique element $r^{\#} \in \prod_{x \in X(R)} Q^r_{cl}(R_x)$, where $Q^r_{cl}(R_x)$ is the classical right ring of quotients of the ring R_x, as follows:

$$r^{\#}_x = \begin{cases} r_x^{-1}, & \text{if } x \in \operatorname{supp} r, \\ 0_x, & \text{otherwise.} \end{cases}$$

Let Q be a subring of $\prod_{x \in X(R)} Q^r_{cl}(R_x)$ generated by the set

$$R \cup \{r^{\#}, \ r\text{-reduced element of } R\}.$$

We shall prove $Q = Q^r_{\pi}(R)$. It is clear that Q is a unital extension of R. As we have noted, we have to test Conditions (i)′ and (ii)′.

Condition (i)′ is obvious, because $r^{\#}$ is the π-regular inverse of the reduced element $r \in R$. Further we shall write r^* instead of $r^{\#}$.

Testing of (ii)′: It is enough to prove the following two conditions:

(ii)$'_1$: for every pair of elements $a, b \in R$, where b is reduced, there exist elements $c, d \in R$ such that d is reduced and $b^*a = cd^*$.

(ii)$'_2$: for every pair of reduced elements $a, b \in R$ there exist elements $c, d \in R$ such that d is reduced and $a^* + b^* = cd^*$.

The first one follows immediately from the π-Ore condition in (3).

For the second one let e and f be the associated idempotents of a and b, respectively (see Definition 1.3). We orthogonalize the sum $a^* + b^*$ by

$$a^* + b^* = a^*(e - ef) + b^*(f - ef) + (a^* + b^*)ef$$

where the summands are mutually orthogonal (we recall, two elements $u, v \in R$ are orthogonal, if $uRv = vRu = (0)$). Hence, if we can find $c, d \in R$ with $(a^* + b^*)ef = cd^*$ we will have

$$a^* + b^* = (e \oplus f + c)[a(e - ef) + b(f - ef) + d]^*$$

(we could suppose, without loss of generality that $c = cef$, $d = def$). Therefore our problem is reduced to the case $e = f$. Let $e = f$. By the π-Ore condition we have $as = bt$ with $\operatorname{ann}_R(s) = \operatorname{ann}_R(b)$ for some elements s and t of R, where s is reduced. Moreover, since a, b and s are reduced elements with the same associated idempotent e and for every $x \in \operatorname{supp} e$ t_x is not a right zero-divisor. Hence t_x is not nilpotent. Therefore t_x is a non-zero divisor in R_x and we can assume that t is a reduced element with associated idempotent e. So we have $a^* = s(bt)^*$ and

$$a^* + b^* = s(bt)^* + eb^* = s(bt)^* + tt^*b^* = (s + t)(bt)^*$$

as desired. Thus the theorem is proved.

Let R be a normal g.p.p. ring such that $Q = Q_\pi^r(R)$. Then we can construct the Pierce sheaf $(S(Q), X(R))$ of the right R-module Q, where $S(Q) = \bigcup_{x \in X(R)} Q_x$, $Q_x = Q/xQ$, $x \in X(R)$. Now we shall consider the ring $Q = Q_\pi^r(R)$ as constructed in the proof of the implication ((2) and (3)) $\Rightarrow$ (1) in Theorem 2.7.

Let $\varphi_x : Q \to Q_{cl}^r(R_x)$ be the natural projection. It is clear that $xQ \subset \subset \ker \varphi_x$. Let $q \in \ker \varphi_x$ and $q = ab$, where $a, b \in R$, b is a reduced element with associated idempotent $e \in B(R)$. Then $0 = q_x = a_x b_x^*$ holds and hence either $a_x = 0_x$ or $b_x = 0$. In the first case we have $a \in xR$ and $q \in xRb^* \subset xQ$; in the second case $e \in x$ and hence $q = ab^* = ab^* e \in eQ \subset xQ$. Hence we can easily verify that $Q_{cl}^r(R_x) \cong Q/xQ$ for all $x \in X(R)$ and so we have proved the following

COROLLARY 2.8 (cf. [10, Lemma 10]). *Let R be a normal g.p.p. ring with $Q = Q_\pi^r(R)$. Then for every $x \in X(R)$, Q_x is the classical right ring of quotients of R_x.*

The ring of right π-regular quotients $Q_\pi^r(R)$ of the ring R is, in particular, a unital extension of R in which every element of R is π-regular, i.e. for every $r \in R$ there exists an element $q \in Q_\pi^r(R)$ with $r^m q r^m = r^m$ for some integer $m > 0$.

Concerning the converse statement we have

THEOREM 2.9. *Let R be a normal ring with $E(R)R = R$. Then the following are equivalent:*

(1) *R is a g.p.p. ring with $Q_\pi^r(R)$,*

(2) *there exists a unital extension Q of R such that*

a) *every element of R is π-regular in Q,*

b) *$E(Q) = E(R)$,*

c) *for every element $q \in Q$ there exists a reduced element $r \in R$ such that $qr \in R$ and $qQ = qrQ$.*

PROOF. (1) $\Rightarrow$ (2). Conditions a) and c) are clear for $Q = Q_\pi^r(R)$. For b) let e be an idempotent of Q. Then $e = ab^*$ for some $a, b \in R$, where b is a reduced element, and by Corollary 2.8 $e_x = 1_x$ for each $x \in \operatorname{supp} e$. In particular, if a' is the reduced part of the element a and e_1, e_2 are the associated idempotents of a' and b', respectively, then we have $e = a'b^*$, and hence $e = e_1 e_2 \in E(R)$, as desired.

(2) $\Rightarrow$ (1). At first we prove that R is g.p.p. ring. Let $r \in R$ and $r^m = r^m q r^m$ for some $q \in Q$ and an integer $m \geq 1$. Then $r^m q, q r^m E(Q) = B(R)$, and $r_R(r^m) = r_R(q r^m) = \operatorname{ann}_R(q r^m)$, $\ell_R(r^m) = \ell_R(r^m q) = \operatorname{ann}_R(r^m q)$. Thus, the ring R is both left and right g.p.p. ring with $E(R)R = R$, hence by definition R is g.p.p. ring.

Now we shall verify that R satisfies the right π-Ore condition. Let $a, b \in \in R$ and b be a reduced element. We can prove that every reduced and

hence every element of R has a π-regular inverse in Q. Therefore, by the assumption for the element $b^*a \in Q$ there exists a reduced element $d' \in R$ such that $b^*ad' \in R$ and $b^*aQ = b^*ad'Q$. Let e and f be the associated idempotents of b and d', respectively. It follows from the last equality that $b^*a = b^*ae = b^*ad'q$ for some $q \in Q$. Since $bb^* = e$, we have $ea = ead'q$, and hence $fea = fead'q = ead'q = ea$, $(e - fe)a = 0$. We put $c = b^*ad'$, $d = e - fe + ed'$. Then $ead' = bc$ and $ad = a(e - fe) + ead' = bc$. Moreover, it is clear that d is a reduced element with $\operatorname{ann}_R d = \operatorname{ann}_R b$. So, we get the right π-Ore condition for R and for R the ring $Q^r_\pi(R)$ exists (Theorem 2.7).

COROLLARY 2.10. *Let R be a normal g.p.p. ring with $Q^r_\pi(R)$. Then $E(Q^2_\pi(R)) = B(R)$ and $Q^r_\pi(R)$ is a normal ring.*

PROOF. By Theorem 2.9 there exists a ring Q with Conditions a), b) and c). Then by the universal property of $Q^r_\pi(R)$ and by Condition a) we conclude that $Q^r_\pi(R) \subseteq Q$. Then by Condition c) we have $Q = Q^r_\pi(R)$. Hence Condition b) gives the desired equality $E(Q^r_\pi(R)) = B(R)$. Now let r be an arbitrary reduced element of R and $e \in B(R)$. Then

$$r^*e = r^*rr^*err^* = r^*rr^*rer^* = er^*rr^*rr^* = er^*.$$

Therefore e is a central idempotent in $Q^r_\pi(R)$.

In the rest of this section we shall specialize our results for rings with identity element. As a key for such specializations we have the following

PROPOSITION 2.11. *Let R be a normal g.p.p. ring with identity. Then R has the classical right ring of quotients $Q^r_{cl}(R)$ if and only if it has the ring $Q^r_\pi(R)$. Moreover, if they exist then $Q^r_{cl}(R) = Q^r_\pi(R)$.*

PROOF. Suppose R has the ring $Q^r_{cl}(R)$. It is clear that if r is a non-zero divisor in R, then r is reduced and $r^{-1} = r^*$. Let r be a reduced element of R with associated idempotent e. Then $r + (1 - e)$ is a non-zero divisor in R and $r^* = e(r + (1 - e))^{-1}$. Therefore $Q^r_{cl}(R)$ is the ring of right π-regular quotients of R: $Q^r_{cl}(R) = Q^r_\pi(R)$.

Conversely, let R be a ring with $Q^r_\pi(R)$. We shall prove that every element $q \in Q^r_\pi(R)$ has the form $q = ab^*$, where b is a non-zero divisor in R. Let $q = cd^*$ for some $c, d \in R$, where d is a reduced element with associated idempotent e. Then putting $a = ce$ and $b = d + (1 - e)$ we have $q = ab^*$. Now it is clear that $Q^r_\pi(R) = Q^r_{cl}(R)$.

From Theorems 2.7 and 2.9 and Proposition 2.11 we immediately have the following

COROLLARY 2.12. *Let R be a normal ring with identity. Then the following are equivalent:*

(1) *R is a g.p.p. ring with $Q^r_{cl}(R)$.*

(2) a) *Every Pierce stalk of R is a right Ore ring in which every zero divisor is nilpotent.*

b) *For every element* $r \in R$, *the set*

$$\{x \in X(R) \mid r_x \text{ is a non-zero divisor}\}$$

is open and closed in $X(R)$.

COROLLARY 2.13. *Let* R *be a normal ring with identity. Then the following are equivalent:*

(1) R *is a g.p.p. ring with* $Q^r_{cl}(R)$.

(2) *There exists an extension* Q *of* R *which contains the same identity as* R *such that*

a) *every element of* R *is* π*-regular in* Q,

b) $E(Q) = E(R)$,

c) *for every element* $q \in Q$, *there exists a non-zero divisor* $r \in R$ *with* $qr \in R$.

(3) R *has the classical ring of right quotients* $Q^r_{cl}(R)$ *such that*

a) *every element of* R *is* π*-regular in* $Q^r_{cl}(R)$,

b) $E(Q^r_{cl}(R)) = E(R)$.

We note that the equivalence of the Conditions (1) and (4) in Corollary 2.1.3 has been proved in [8, Theorem 2].

3. Normal g.p.p. rings with Köthe radical

We recall that a ring R is said to be a *ring with Köthe radical* $K(R)$ if $K(R)$ is a nil ideal of R containing all one-sided nil ideals of R. It is a well-known open question whether every ring has Köthe radical. We also recall that a ring R is called a *local ring* if the ring $R/J(R)$ is a division ring, where $J(R)$ denotes the Jacobson radical of R.

From Lemma 2.1 and from the proof of the implication $(2) \Rightarrow (3)$ in Theorem 2.7 we have

PROPOSITION 3.1. *Let* R *be a normal ring. Then* R *is a g.p.p. ring if and only if the following conditions are satisfied:*

1) *in each Pierce stalk* R_x *of* R *every zero divisor is nilpotent,*

2) *for every element* $r \in R$ *the set*

$$\{x \in X(R),\ r_x \text{ is non-zero divisor in } R_x\}$$

is open compact in $X(R)$ *and* $\operatorname{supp} r \subseteqq \operatorname{supp} f$ *for an idempotent* $f \in B(R)$.

An addition needed for the main result of this section and also having its own interest is the following.

PROPOSITION 3.2. *Let R be a right Ore ring. Then the following are equivalent:*

1) *for an arbitrary element $r \in R$ either r is non-zero divisor, or $r+s$ is non-zero divisor for all non-zero divisor $s \in R$,*

2) $Q^r_{cl}(R)$ *is a local ring.*

PROOF. 1) $\Rightarrow$ 2). Let $q = rs^{-1} \in Q$ be a non-invertable element of Q, where $Q = Q^r_{cl}(R)$ and $r, s \in R$. Then r is zero divisor, and hence by the assumption $r+s$ is a non-zero divisor and $1+rs^{-1} = (r+s)s^{-1}$ is invertable. Now we verify that

$$J(Q) = \{q \in Q,\ q \text{ is not invertable}\}.$$

The inclusion $\subseteq$ is obvious. Let q be a non-invertable element. Now q is a zero divisor in Q. Let p be an arbitrary element of Q. If q is a right zero divisor (the left case is similar) then qp is a right zero divisor, and hence, as we have proved, $1+qp$ is invertable. Thus qQ as a quasi-regular right ideal must be contained in $J(Q)$, in particular, $q \in J(Q)$. Therefore $Q/J(Q)$ is a division ring and Q is a local ring.

2) $\Rightarrow$ 1) Let r be a right zero divisor in R, and s be a non-zero divisor in R. We have to verify that $r+s$ is a non-zero divisor. Since $q = rs^{-1}$ is a non-invertable element of Q by assumption $q \in J(Q)$ and $1+rs^{-1}$ is invertable. Therefore $r+s = (1+rs^{-1})s$ is invertable in Q and $r+s$ is a non-zero divisor in R.

THEOREM 3.3. *Let R be a normal ring.*

(A) *The following conditions are equivalent:*

a) *R is a g.p.p. ring with Köthe radical,*

b) *in each Pierce stalk R_x every zero divisor is nilpotent, and the sum of nilpotent elements in R_x is nilpotent and for every element $r \in R$ the set*

$$\{x \in X(R),\ r_x \text{ is non-zero divisor}\}$$

is open and compact and $\operatorname{supp} r \subseteq \operatorname{supp} f$ *for some $f \in B(R)$.*

(B) *Moreover, if R is a g.p.p. ring with $Q^r_\pi(R)$ the above conditions are equivalent with the following:*

c) *every Pierce stalk Q_x of the ring $Q^r_\pi(R)$ on $x \in X(R)$ is a local ring.*

PROOF. At first we prove the following

LEMMA 3.4. *Let R be a normal g.p.p. ring. Then R has Köthe radical if and only if the set of all nilpotent elements of R is an ideal of R.*

PROOF of the Lemma. The necessity is obvious. Let R be normal g.p.p. ring with Köthe radical K and let a be a nilpotent element of R. Then for every element $b \in R$ ab is nilpotent, because $(ab)_x = a_x b_x$ is a zero divisor,

hence the reduced part of the element ab is zero. Therefore aR is a nil right ideal. In particular, $a \in K$ and K is an ideal of all nilpotent elements of R.

Now we continue the proof of Theorem 3.3.

a) $\Rightarrow$ b): Let a_x, b_x be nilpotent elements of R_x. Then for some $e \in B(R)$ with $e_x = 1_x$, the elements ae and be are nilpotent. So, by the assumption and by Lemma 3.4 $ae + be$ is nilpotent, and hence $a_x + b_x$ is nilpotent. The rest of Condition b) follows from Proposition 3.1.

b) $\Rightarrow$ a). It is enough to verify that the sum of nilpotent elements of R is nilpotent and this can be done by the standard Pierce sheave argument (see the proof of Theorem 2.7 (2) $\Rightarrow$ (3)).

(B) b) $\Rightarrow$ c). Let R be a ring with $Q^r_\pi(R)$ and let Condition b) in (A) be satisfied. By Corollary 2.8 $Q_x = Q^r_{cl}(R_x)$ for all $x \in X(R)$. Moreover, R_x satisfies Condition 1) of Proposition 3.2, because if $r_x + s_x$ and r_x are nilpotent, then s_x is nilpotent, too. Therefore by this proposition Q_x is a local ring.

c) $\Rightarrow$ b): Let $Q_x = Q^r_{cl}(R_x)$ be a local ring, and $a_x, b_x \in R_x$ be nilpotent. If $a_x + b_x$ is not nilpotent, then it is non-zero divisor and hence by Proposition 3.3 $a_x = (a_x + b_x) - b_x$ is non-zero divisor. This is, however, impossible and so $a_x + b_x$ is nilpotent.

COROLLARY 3.4. *Let R be a normal g.p.p. ring with classical right ring of quotients Q. Then every Pierce stalk of the ring Q is local ring if and only if R has Köthe radical.*

For the proof we note that $B(Q) = B(R)$.

With respect to Corollary 3.4 we have the following

CONJECTURE. *Let Q be a classical right ring of quotients of a normal g.p.p. ring with identity. Then every Pierce stalk of the ring Q is a local ring.*

4. Normal g.p.p. ring with π-regular ring of right π-regular quotients

In this section we shall give a characterization for the rings in the title of this section.

We recall that if Q is the classical right ring of quotients of a ring R then R is called a *right order* in Q.

THEOREM 4.1. *Let R be a normal ring. Then the following are equivalent:*

1) *R is a g.p.p. ring with $Q^r_\pi(R)$ which is π-regular,*

2) *R is a g.p.p. ring with $Q^r_\pi(R)$ which is g.p.p. ring,*

3) a) *every Pierce stalk R_x of R is a right order in a π-regular ring and every zero-divisor in R_x is nilpotent;*

b) *for every element* $r \in R$ *the set*

$$\{x \in X(R) \mid r_x \text{ is non-zero divisor}\}$$

is open and compact and $\operatorname{supp} r \subseteqq \operatorname{supp} f$ *for an* $f \in B(R)$.

4) *There exists a unital extension* Q *of* R *such that*

c) Q *is* π*-regular,*

d) $E(Q) = B(R)$,

e) *for every* $q \in Q$ *there exists a reduced element* $r \in R$ *such that* $qr \in R$ *and* $qQ = qrQ$.

PROOF. 1) $\Rightarrow$ 2) is clear, since any π-regular (normal) ring is a g.p.p. ring.

2) $\Rightarrow$ 3): By Corollary 2.8 for verifying Condition a) in 3) it is enough to prove that every Pierce stalk $Q_x = Q^r_{cl}(R_x)$ of $Q = Q^r_\pi(R)$ is a π-regular ring. Since by the assumption Q is g.p.p. ring, in Q_x every non-invertable element is nilpotent. Therefore Q_x is π-regular, as desired. Condition b) is obvious.

3) $\Rightarrow$ 4): By Theorem 2.7 R is g.p.p. ring with $Q^r_\pi(R)$ and by Corollary 2.8 every stalk of $Q^r_\pi(R)$ is π-regular. Therefore, by the standard Pierce sheave argument we can prove that $Q^r_\pi(R)$ is π-regular (see [2, Proposition 3.3]).

4) $\Rightarrow$ 1): By Theorem 2.9 R is a g.p.p. ring with $Q^r_\pi(R)$. Moreover, one can prove that $Q = Q^r_\pi(R)$ (see the proof of Corollary 2.10), and hence $Q^r_\pi(R)$ is π-regular.

LEMMA 4.2. *Let* R *be an indecomposable normal ring with identity in which every zero divisor is nilpotent. Assume that* R *has* $Q = Q^r_{cl}(R)$. *Then the following conditions are equivalent:*

1) Q *is local ring whose Jacobson radical is nil,*

2) Q *is a* π*-regular ring,*

3) *if* $a, b \in R$ *and* a *is nilpotent, and* b *is non-zero divisor, then* ab^{-1} *is nilpotent.*

PROOF. Obvious.

From Theorem 4.1, Proposition 2.11 and Lemma 4.2 we get the following

COROLLARY 4.3 (cf. [10, Theorem 1]). *Let* R *be a normal ring with identity. Then the following are equivalent:*

(1) R *is a right order in a* π*-regular ring* Q *and* $E(Q) = B(R)$.

(2) R *is right Ore ring and both* R *and* $Q^r_{cl}(R)$ *are g.p.p. rings.*

(3) R *is a right order in* Q *and for any* $x \in X(R)$ *the Pierce stalks of* Q *on* $X(R)$ *are local rings whose Jacobson radicals are nil.*

(4) a) *For any* $x \in X(R)$ *the Pierce stalk* R_x *is a right order in a* π*-regular ring, and every zero-divisor in* R_x *is nilpotent,*

b) *for any* $r \in R$ *the set*

$$\{x \in X(R),\ r_x \text{ is non-zero divisor}\}$$

is open and closed,

(5) c) *for any $r \in R$ there is a positive integer n such that for any $m \geqq n$,* $\operatorname{supp} r^m = \operatorname{supp} r^n$ *is open and closed,*

d) *for any $x \in X(R)$ the Pierce stalk R_x is a right order in a ring $Q^{(x)}$ such that any zero divisor $r \in R_x$ and any $q \in Q^{(x)}$ rq is nilpotent.*

Acknowledgement. The author wishes to express his indebtedness and gratitude to Prof. L. A. Bokut, for his permanent attention and to the Referee for helpful suggestions.

REFERENCES

[1] Burgess, W. D. and Stephenson, W., Pierce sheaves of noncommutative rings, *Comm. Algebra* **4** (1976), 51–75. *MR* **53** #8122

[2] Burgess, W. D. and Stephenson, W., Rings all of whose Pierce stalks are local, *Canad. Math. Bull.* **22** (1979), 159–164. *MR* **80g**: 16018

[3] Dauns, J. and Hofmann, K. H., The representation of biregular rings by sheaves, *Math. Z.* **91** (1966), 103–123. *MR* **32** #4151

[4] Gonchigdorzh, R., A ring of regular quotients of a reduced ring, *Algebra i Logika* **26** (1987), 150–164 (in Russian). *MR* **89m**: 16008a

[5] Gonchigdorzh, R., Reduced p.p. rings and semihereditary and hereditary rings, *Algebra i Logika* (in Russian) (to appear).

[6] Hirano, Y., On generalized p.p. rings, *Math. J. Okayama Univ.* **25** (1983), 7–11. *MR* **84g**: 13020

[7] Hirano, Y. and Tominaga, H., Rings decomposed into direct sums of nil rings and certain reduced rings, *Math. J. Okayama Univ.* **27** (1985), 35–38. *MR* **87d**: 16033

[8] Ôhori, M., On non-commutative generalized p.p. rings, *Math. J. Okayama Univ.* **26** (1984), 157–167. *MR* **86d**: 16018

[9] Ôhori, M., On strongly π-regular rings and periodic rings, *Math. J. Okayama Univ.* **27** (1985), 49–52. *MR* **87d**: 16034

[10] Ôhori, M., Some studies on generalized p.p. rings and hereditary rings, *Math. J. Okayama Univ.* **27** (1985), 53–70. *MR* **87f**: 16023

[11] Pierce, R. S., *Modules over commutative regular rings*, Mem. Amer. Math. Soc., No. 70, Amer. Math. Soc., Providence, 1967. *MR* **36** #151

[12] Johnstone, P. T., *Stone spaces*, Cambridge studies in advanced mathematics **3**, Cambridge University Press, Cambridge – New York, 1982. *MR* **85f**: 54002

(Received May 6, 1987)

INSTITUTE OF MATHEMATICS
MONGOLIAN ACADEMY OF SCIENCES
P.O.BOX 112
ULAN-BATOR 51
MONGOLIAN PEOPLE'S REPUBLIC

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 17–61

DIALECTICAL LOGIC: THE PROCESS CALCULUS

παλιντονος αρμονιη — παλιντροπος αρμονιη

R. E. KENT

Abstract

Dialectical logic is the logic of dialectical processes. The goal of dialectical logic is to reveal the dynamical notions inherent in logical computational systems. The fundamental notions of *proposition* and *truth-value* in standard logic are subsumed by the notions of *process* and *flow* in dialectical logic. Standard logic motivates the core sequential aspect of dialectical logic. Horn-clause logic requires types and nonsymmetry and also motivates the parallel aspect of dialectical logic. The process logics of Milner and Hoare reveal the internal/external aspects of dialectical logic. The sequential internal aspect of dialectical logic should be viewed as a typed or distributed version of Girard's linear logic with nonsymmetric tensor. The simplest version of dialectical logic is inherently intuitionistic. However, by following Glivenko's approach in standard logic using double negation closure, we can define a classical version of dialectical logic.

Introduction

Abstract objective knowledge, such as general science and philosophy, originated in the fifth and sixth centuries B.C. in the thought, teachings and writings of the preSocratic Greek philosophers. The aim of the preSocratics was to give a nonmythological account of the origin of the world (**kosmos**), and to rationally explain its motion. By far the most common explanation given by the preSocratics for the origin and motion of the kosmos was in terms of pairs of opposing tendencies, such as the *hot* and the *cold*, the *wet* and the *dry*, *love* and *strife*, etc. In fact, the notion of complementary pairs of opposing tendencies has occurred throughout the history of ideas. Ancient examples of opposing tendencies occur not only in preSocratic Greek philosophy, but also in naturalistic Chinese philosophy, as the dualistic concept of *yin* and *yang*; and in Indian Hindu philosophy, as *Brahma* the creator and *Shiva* the destroyer with *Vishnu* the preserver.

For the preSocratics, who were postmythological but prelogical, the components of such opposed pairs were neither properties nor objects, but motive forces. The dynamics in this world-view is obvious. Unfortunately, much of this dynamical world-view was lost to the history of ideas when logic was

1991 *Mathematics Subject Classification.* Primary 03F99.

Key words and phrases. Dialectic, contradiction, adjunction, flow, constraint, structure, interaction, process, object, calculus, sequent, polar, complete, sound, Heyting, linear logic, negation, residuation, orthogonality, tensor, classical, intuitionistic, Boolean.

Akadémiai Kiadó, Budapest

conceived as a study of static notions. A central theme of this paper is that much of this dynamical world-view needs to be re-revealed, re-developed, and extended, in order to comprehend modern logical computational systems. A modern theory of dialectics offers the appropriate conceptual framework for doing this; it takes the notion of opposing tendencies as its central concept, and calls it *dialectical contradiction.* This modern dialectical theory still retains the motive force interpretation for the components (aspects) of dialectical contradictions: dialectical contradictions specify *dialectical motion*, where motion is not mere physical motion, but any change whatsoever; motion is synonymous with transformation. The distinction between the concepts of dialectical contradiction and dialectical motion, two fundamental notions of dialectics, is itself dialectical, the *potential* aspect and the *actual* aspect. These two concepts occur in ancient and modern interpretations of the fragments of Heraclitus, the most dialectically oriented preSocratic [Hussey], and are contained here in the subtitle: $\pi\alpha\lambda\iota\nu\tau o\nu o\varsigma$ $\alpha\rho\mu o\nu\iota\eta$ — $\pi\alpha\lambda\iota\nu\tau\rho o\pi o\varsigma$ $\alpha\rho\mu o\nu\iota\eta$ (**palintonos harmonie — palintropos harmonie**); (crudely) polar tension structure — polar turning structure; the "tension" interpretation — the "oscillation" interpretation, of Heraclitus; or for us, dialectical contradiction — dialectical motion.

The history of dialectics is replete with intuitively suggestive, but ill-defined and non-rigorous, ideas and examples [Bernow, Piccone]. If the dialectical point of view is to be useful as a human conceptual structure, its objective aspect must have a rigorous foundation. The notion of dialectical contradiction is monistically objectified [Lawvere] by the mathematical idea of *adjunction.* Since adjoint pairs are (one of) the most important concepts of category theory, this point-of-view is summarized by the statement: Category Theory ≡ Objective Dialectics. The notion of dialectical contradiction is pluralistically objectified [Kent87] by the mathematical idea of *dialectical base.* In objective dialectics, since dialectical contradictions are represented by adjunctions, systems of dialectical contradictions are represented by diagrams in the unbounded category (**to apeiron**) whose morphisms are adjoint pairs of functors. Such a diagram, whose component categories usually have certain completeness properties, is called a *dialectical base* of preorders. From a static, non-dynamic, non-dialectical point-of-view, this has also been called an indexed category [Hyland]. Within the notion of dialectical contradiction the distinction between the concepts of adjunctions and dialectical bases is dialectical, the *one-many* dialectic.

The notion of dialectical motion can be specified [Kent87] by the mathematical idea of *dialectical system*, or parallel pair of distributed *terms.* Dialectical systems have the following essential aspects: [**ancient**] they are based upon contradictions or opposing tendencies; they define motion, flow or development; [**modern**] they contain internally interacting and combining objects or entities in dialectical motion; and they specify the reproduction or renewal of such entities, where reproduction is equilibrium of dialectical motion. Dialectical systems are the "motors of nature" specifying the di-

alectical motion of structured entities, and a dialectical base provides the "motive power" for this motion. The notion of dialectical motion can be realized by the mathematical idea of *dialectical flow*, which is the oscillation (alternation-composition) of inverse flow along one term and direct flow along the other term. Direct and inverse flow are suitably generalized *Kan extensions* which make use of a dialectical base. Dialectical systems specify dialectical flow, and dialectical flow is the realization of dialectical systems; the *specification-realization* dialectic.

It has been known for some time now [Lawvere] that logic is dialectical in nature, but the full force of its dialectical structure has only recently [Girard, Kent88] been discussed. Dialectical ideas, not only come chronologically and historically before logical ideas, but also come conceptually before them as well. The theory and practice of computer science and dynamic systems contain many dialectical contradictions. Two of the most important of these, the *flow* dialectic and the *constraint* dialectic, constitute the proper study of dialectical logic [Kent88]; whereas a third, the *part-whole* dialectic, is important in its standard aspect [Kent89]. Dialectical logic is the logic of dialectical processes. It invests the dynamical view of systems theory with the fundamental ideas of category theory; but in turn, it gives these categorical notions that dynamical view. Dialectical logic provides a unified semantics for both the object paradigm and the process paradigm of programming-in-the-large. By subsuming process logic [Milner, Hoare78] along with clause logic, it allows the specification of strongly-typed parallel logic programs. In dialectical logic aspects of the process paradigm are modelled as a flow dialectic, whereas aspects of the object paradigm are modelled as a constraint dialectic orthogonal to flow. The flow (or *product-implication*) dialectic is the internal aspect of dialectical logic, whereas the constraint dialectic is its external aspect.

Dialectical logic is based upon the two interdependent concepts of structure and dialecticality. Dialecticality is built out of the aspects of dialectical tension and dialectical flow, as mentioned above. Structure is concentrated in the compositionality of monoids and comonoids (this includes the grand unification principle [Manes] that "composition determines semantics"), and in the type-summability of orthogonal terms (the object calculus, discussed below). Structure occurs peripherally in the interactions of limits, the combinations of colimits, and the reproduction of fixpoints. The structurality of limits and colimits, being special Kan extensions, has obvious dialecticality. This is but one indication of the interdependence of structure and dialectics; other indications are the simple facts that monoids have associated adjoint pairs, and adjoint pairs compose into monoids and comonoids. Parsimonious use of (1) abstract monoidal concepts for modelling "construction", "composition" and "interaction", along with (2) adjointness notions for modelling "dialectical flow" (such as "predicate transformation") has great potential in the computational and system sciences.

Dialectical logic is an extension of standard logic. The extension of

propositional calculus is called the *process calculus*; the extension of predicate calculus is called the *object calculus*. In this paper we are mainly concerned with the process calculus; its intuitionistic and classical semantics, and its classical axiomatics. In a succeeding paper [Kent88] we will be concerned chiefly with the object calculus. In order that readers may begin to explore the fascinating possibilities of dialectics, I have included in the appendix to this paper an introduction to this object aspect of dialectics.

1. Preliminaries

Dialectical laws. The "laws of dialectics" are laws of logic. The most fundamental dialectical law, the law of *the interpenetration of opposites*, is represented in general by adjoint pairs of functors or monotonic functions, and in particular by the flow dialectic (tensor product — tensor implication adjointness). As a special case of this, the dialectical law of *the negation of the negation* is represented in general as a self-adjoint functor or monotonic function, and in particular by tensor negation. Here we discuss the general case. The paper as a whole is a discussion of the particular case.

Two opposed monotonic functions $\langle B, \leqq_B \rangle \xrightarrow{f} \langle A, \leqq_A \rangle$ and $\langle B, \leqq_B \rangle \xleftarrow{g} \langle A, \leqq_A \rangle$ between preorders form an *adjoint pair*, denoted $f \dashv g$, when they satisfy the equivalence $f(b) \leqq_A a$ iff $b \leqq_B g(a)$. This equivalence can be interpreted as the "polar-tension structure" of the preSocratic Greek philosopher Heraclitus [Hussey], and in Greek is rendered $\pi\alpha\lambda\iota\nu\tau o\nu o\varsigma\ \alpha\rho\mu o\nu\iota\eta$. The fact that $f \dashv g$ is an adjoint pair is equivalently defined by the "unit" inequality $B \leqq f \cdot g$ and the "counit" inequality $g \cdot f \leqq A$. The composite monotonic functions $\langle B, \leqq_B \rangle \xrightarrow{f \cdot g} \langle B, \leqq_B \rangle$ and $\langle A, \leqq_A \rangle \xrightarrow{g \cdot f} \langle A, \leqq_A \rangle$ are closure and interior operators, respectively. A *closure operator* $\langle B, \leqq_B \rangle \xrightarrow{k} \langle B, \leqq_B \rangle$ is a monotonic endofunction which is "increasing" $B \leqq k$ and "idempotent" $k \cdot k = k$. Dually, an *interior* (or *kernel*) *operator* $\langle A, \leqq_A \rangle \xrightarrow{j} \langle A, \leqq_A \rangle$ is a monotonic endofunction which is "decreasing" $A \geqq j$ and "idempotent" $j \cdot j = j$. An adjoint pair $f \dashv g$ is a *reflective pair* when the counit is an equality $g \cdot f = A$, stating that the interior operator $g \cdot f$ is an identity. So an adjoint pair $f \dashv g$ is a reflective pair iff f is a surjective monotonic function iff g is an injective monotonic function. An adjoint pair $f \dashv g$ is a *coreflective pair* when the unit is an equality $B = f \cdot g$, stating that the closure operator $f \cdot g$ is an identity. So an adjoint pair $f \dashv g$ is a coreflective pair iff f is an injective monotonic function iff g is a surjective monotonic function.

The corestriction $\langle B, \leqq_B \rangle \xrightarrow{(\ )^\bullet_k} \langle k(B), \leqq_{k(B)} \rangle$ of a closure operator k to its image $k(B) \stackrel{\text{df}}{=} \{k(b) \mid b \in B\}$ of *k-closed elements* of B forms a reflective pair $(\ )^\bullet_k \dashv \text{Inc}$ with the inclusion $\langle k(B), \leqq_{k(B)} \rangle \xrightarrow{\text{Inc}} \langle B, \leqq_B \rangle$. The corestriction $\langle A, \leqq_A \rangle \xrightarrow{(\ )^\circ_j} \langle j(A), \leqq_{j(A)} \rangle$ of an interior operator j to its image $j(A) \stackrel{\text{df}}{=} \{j(a) \mid$

$| a \in A\}$ of *j-open elements* of A forms a coreflective pair $\mathrm{Inc} \dashv (\ )^\circ_j$ with the inclusion $\langle j(A), \leqq_{j(A)} \rangle \xrightarrow{\mathrm{Inc}} \langle A, \leqq_A \rangle$. So for any adjoint pair $f \dashv g$, the subpreorders of $f \cdot g$-closed elements $B^\bullet \subseteq B$ and $g \cdot f$-open elements $A^\circ \subseteq A$ participate themselves in the special adjunctions $(\ )^\bullet \dashv \mathrm{Inc}$ and $\mathrm{Inc} \dashv (\ )^\circ$ of reflective and coreflective pairs, respectively. The restriction of the adjoint pair to closed/open elements forms an inverse pair of monotonic functions, making B-closed elements isomorphic to A-open elements $B^\bullet \cong A^\circ$. The adjoint pair, the closed element reflection, the open element coreflection, and the inverse pair, form a commuting square of dialectical contradictions. For a reflexive pair $f \dashv g$, all elements of A are open $A^\circ = A$, and hence A is isomorphic to the B-closed elements $B^\bullet \cong A$. Any reflective pair $f \dashv g$ is equivalent to the $(\ )^\bullet \dashv \mathrm{Inc}_{B^\bullet}$ reflective pair which factors the closure operator $f \cdot g$ through its image $B^\bullet$. For a coreflective pair $f \dashv g$, all elements of B are closed $B^\bullet = B$, and hence B is isomorphic to the A-open elements $B \cong A^\circ$. Any coreflective pair $f \dashv g$ is equivalent to the $\mathrm{Inc}_{A^\circ} \dashv (\ )^\circ$ coreflective pair which factors the interior operator $g \cdot f$ through its image A°. So any inverse pair is an adjoint pair with the identity orderings, and any adjoint pair determines an inverse pair. Adjointness is a kind of generalized inverseness (another related kind of generalized inverseness is the notion of orthogonality defined below).

The special case of self-adjointness, where $f = g^{\mathrm{op}}$ and $A = B^{\mathrm{op}}$, defines the notion of "negation". When a monotonic function $\langle A, \leqq \rangle \xrightarrow{f} \langle A, \leqq \rangle^{\mathrm{op}}$ is self-adjoint $f \dashv f^{\mathrm{op}}$ it is called a *negation*. The polar-tension structure is the equivalence $a \leqq f(a')$ iff $a' \leqq f(a)$, and A-closed elements and A^{op}-open elements coincide, with dialecticality expressed as duality $A^\bullet \cong (A^{\mathrm{op}})^\circ = (A^\bullet)^{\mathrm{op}}$. So restricting f to the f^2-closed elements of A makes f into an *involution*: "idempotent" $f^2(a) = a$, "monotonic" if $a \leqq b$ then $f(b) \leqq f(a)$, and satisfying $f(a \vee b) = f(a) \wedge f(b)$ (a De Morgan's law) and $f(\bot) = \top$ when the joins exist.

Biposets. A *biposet* is another name for an ordered category; that is, a category $\mathbf{P} = \langle \mathbf{P}, \preceq, \circ, \mathrm{Id} \rangle$ whose homsets are posets and whose composition is monotonic on left and right. We prefer to view biposets as vertical structures, preorders with a tensor product, rather than as horizontal structures, ordered categories.

In more detail, a biposet $\mathbf{P}$ consists of the following data and axioms. There is a collection of $\mathbf{P}$-objects $x, y, z, \cdots$ called *types*, and a collection of $\mathbf{P}$-arrows $r, s, t, \cdots$ called *terms*. Terms could also be called "preprocesses", since processes (which are discussed in [Kent88]) are terms which satisfy certain constraints or closure conditions. Each term r has a unique source type y and a unique target type x, denoted by the relational notation $y \xrightarrow{r} x$. The collection of terms from source type y to target type x is ordered by a binary relation $\preceq_{y,x}$ called *term entailment*, which is transitive, if $r \preceq s$ and $s \preceq t$ then $r \preceq t$, reflexive $r \preceq r$, and antisymmetric, $r \equiv s$ implies $r =$

$= s$, where $r \equiv s$ means $r \preceq s$ and $s \preceq r$. Dialectical logic entailment $\preceq_{y,x}$ between terms generalizes standard logic entailment $\vdash$ between propositions. For any two terms $z \xrightarrow{s} y$ and $y \xrightarrow{r} x$ with matching types (target type of s = source type of r) there is a composite term $z \xrightarrow{s \circ r} x$, where $\circ$ is a binary operation called *tensor product*, which is associative $t \circ (s \circ r) = (t \circ s) \circ r$, and monotonic on left and right, $s \preceq s'$ and $r \preceq r'$ imply $(s \circ r) \preceq (s' \circ r')$. Tensor product allows each term $y \xrightarrow{r} x$ to specify a *right direct flow* $\mathbf{P}[z,y] \xrightarrow{r \circ} \mathbf{P}[z,x]$ and a *left direct flow* $\mathbf{P}[x,z] \xrightarrow{r \circ} \mathbf{P}[y,z]$ for each type z. Any type x is a term $x \xrightarrow{x} x$, which is an identity, $r \circ x = r$ and $x \circ s = s$. A biposet with one object (universal type) is called a *monoidal poset*. For each $\mathbf{P}$-type x, the collection $\mathbf{P}[x,x]$ of endoterms at x is a monoidal poset. If $\mathbf{P}$ is a biposet, then the op-dual or opposite biposet $\mathbf{P}^{\mathrm{op}}$ is the opposite category with the same homset order as $\mathbf{P}$, and the co-dual biposet $\mathbf{P}^{\mathrm{co}}$ is (the same category) $\mathbf{P}$ with the opposite homset order. A *morphism of biposets* $\mathbf{P} \xrightarrow{H} \mathbf{Q}$ is a functor which preserves homset order. Any Heyting algebra is a biposet, where tensor product coincides with lattice meet $s \circ r = s \wedge r$. The category **Rel** of sets and (binary) relations is a biposet, where tensor product is relational composition $S \circ R \stackrel{\mathrm{df}}{=} \{(z,x) \mid \exists_{y \in Y} (z,y) \in S \text{ and } (y,x) \in R\}$. A bimodule $\mathcal{Y} \xrightarrow{R} \mathcal{X}$ between two preorders $\mathcal{Y} = \langle Y, \leqq_Y \rangle$ and $\mathcal{X} = \langle X, \leqq_X \rangle$ is a monotonic function $\mathcal{Y}^{\mathrm{op}} \times \mathcal{X} \xrightarrow{R} 2$. The category **Bim** of preorders and preorder bimodules (bimodules $\mathcal{Y} \xrightarrow{R} \mathcal{X}$ are in bijection with closed-above subsets $R \subseteq \mathcal{Y}^{\mathrm{op}} \times \mathcal{X}$) is a biposet, where tensor product is again relational composition $S \circ R \stackrel{\mathrm{df}}{=} \{(z,x) \mid \exists_{y \in Y} (z,y) \in S \text{ and } (y,x) \in R\}$. Given an alphabet A, the category of formal A-languages $\mathcal{P}(A^*)$ is a biposet; whose arrows are formal languages, whose composition is language concatenation, and whose identity is singleton empty string $\{\varepsilon\}$.

Given two types y and x in a biposet $\mathbf{P}$, two opposed terms (terms oppositely directed) $x \xrightarrow{s} y$ and $x \xleftarrow{r} y$ are *semi-orthogonal* at x, denoted $s \perp_x r$, when $s \circ r \prec_{x,x} x$. Semi-orthogonality is a nonsymmetric notion. By combining semi-orthogonality at source and target we get a symmetric notion: two opposed terms $y \xrightarrow{r} x$ and $y \xleftarrow{s} x$ form an orthogonal pair of terms or an *orthoterm*, denoted by $y \xrightarrow{r \perp s} x$, when they satisfy semi-orthogonality at y and semi-orthogonality at x; that is, $r \perp s$ iff $(r \circ s \preceq y \text{ and } s \circ r \preceq x)$. In this case, we say that r is *orthogonal* to s. Orthoterms axiomatize "ring-structured $\mathbf{P}$-terms". Orthoterms compose in the obvious way: $(s \perp s') \circ (r \perp r') = (s \circ r) \perp (r' \circ s')$, and $(x \perp x)$ is the identity orthoterm at x. The homset order on orthoterms is defined by: $(p \perp q) \preceq (r \perp s)$ when $p \preceq r$ and $q \succeq s$. So each biposet $\mathbf{P}$ has an associated *orthoterm category* $\mathbf{P}^{\perp}$, whose objects are $\mathbf{P}$-types and whose arrows are $\mathbf{P}$-orthoterms. There are two projection functors $\mathbf{P}^{\mathrm{op}} \xleftarrow{\partial_0} \mathbf{P}^{\perp} \xrightarrow{\partial_1} \mathbf{P}$, whose product pairing functor is the inclusion $\mathbf{P}^{\perp} \xrightarrow{\mathrm{Inc}} \mathbf{P}^{\mathrm{op}} \times$

$\times \mathbf{P}$. Let $\perp(r)$ denote the collection of all terms opposed and orthogonal to r; $\perp(r) \stackrel{\text{df}}{=} \{x \stackrel{s}{\rightharpoonup} y \mid r \perp s\}$. Then $\perp(r)$ is a closed-below subset of $\mathbf{P}[x, y]$. In defining the phase semantics for linear logic, Girard *implicitly* uses the notion of orthogonality with respect to a single subset of "antiphases" $\perp$. Since orthogonality is defined with respect to types (identity endoterms) $x, y, z, \cdots$, Girard's set of antiphases $\perp$ corresponds to any arbitrary $\mathbf{P}$-type. Orthogonality of terms in biposets for dialectical logic generalizes disjointness of elements in Heyting algebras for standard logic.

A monoid $\mathbf{M}$ is *symmetric* (or *commutative*) when its tensor product is commutative: $s \circ r = r \circ s$. More generally, a biposet $\mathbf{P}$ is *quasisymmetric* or *orthogonally balanced* when $s \perp_x r$ implies $r \perp_y s$ for all $\mathbf{P}$-types y and x and all opposed pairs of $\mathbf{P}$-terms $y \stackrel{r}{\rightharpoonup} x$ and $y \stackrel{s}{\leftharpoondown} x$. Obviously, these implications can be replaced by logical equivalences. Quasisymmetry asserts that semi-orthogonality is equivalent to orthogonality: $r \perp s$ iff $s \perp_x r$ iff $r \perp_y s$. A symmetric monoidal poset (ordered commutative monoid) is quasisymmetric as a one object biposet.

Internal dialectics. For any opposed pair of ordinary relations $Y \stackrel{R}{\rightharpoonup} X$ versus $Y \stackrel{S}{\leftharpoondown} X$ the "unit inequality" $Y \subseteq R \circ S$ and the "counit inequality" $S \circ R \subseteq X$ *taken together* are equivalent to the facts that R is the graph $R = \mathrm{y}^1(f) = \{(y, f(y)) \mid y \in Y\}$ of a function $Y \stackrel{f}{\rightarrow} X$ and that S is the transpose $S = R^{\mathrm{op}} = \mathrm{y}^1(f)^{\mathrm{op}} = \mathrm{y}^0(f) = \{(f(y), y) \mid y \in Y\}$. On the other hand, the graph $Y \stackrel{\mathrm{y}^1(f)}{\rightharpoonup} X$ of any function $Y \stackrel{f}{\rightarrow} X$ and its transpose $\mathrm{y}^0(f) = (\mathrm{y}^1(f))^{\mathrm{op}}$ satisfy the unit and counit inequalities. So these conditions describe functionality in the biposet **Rel**. For any opposed pair of preorder bimodules $\mathcal{Y} \stackrel{R}{\rightharpoonup} \mathcal{X}$ versus $\mathcal{Y} \stackrel{S}{\leftharpoondown} \mathcal{X}$ where $\mathcal{X}$ is a complete lattice, the "unit inequality" $\mathcal{Y} \subseteq R \circ S$ and the "counit inequality" $S \circ R \subseteq \mathcal{X}$ *taken together* are equivalent to the facts that R is the graph $R = \mathrm{y}^1(f) = \{(y, x) \mid f(y) \leqq_X x\}$ of a monotonic function $\mathcal{Y} \stackrel{f}{\rightarrow} \mathcal{X}$ where f is given by $f(y) = \bigwedge\{x \in X \mid yRx\}$, and that S is the transposed graph of f's order-theoretic involution $S = (\mathrm{y}^1(f^{\propto}))^{\mathrm{op}} = \mathrm{y}^0(f) = \{(x, y) \mid x \leqq_X f(y)\}$ with f given by $f(y) = \bigvee\{x \in X \mid xSy\}$. On the other hand, the graph $\mathcal{Y} \stackrel{\mathrm{y}^1(f)}{\rightharpoonup} \mathcal{X}$ of any monotonic function $\mathcal{Y} \stackrel{f}{\rightarrow} \mathcal{X}$ and its transpose $\mathrm{y}^0(f) = (\mathrm{y}^1(f^{\propto}))^{\mathrm{op}}$ satisfy the unit and counit inequalities. So these conditions describe functionality in a part of the biposet **Bim**. In the general case, when $\mathcal{X}$ is not necessarily complete, the "unit inequality" $Y \subseteq R \circ S$ and the "counit inequality" $S \circ R \subseteq X$ *taken together* are equivalent to the facts that R is the tensor implication (**Bim** is a Heyting category) $R = S\text{-}\backslash\, \mathcal{X} = \{(y, x) \mid (\forall x')$ if $x' S y$ then $x' \leqq_X x\}$ and that S is the implication $S = \mathcal{X} /\text{-} R = \{(x, y) \mid (\forall x')$ if yRx' then $x \leqq_X x'\}$. So these conditions describe a potential functionality in the entire biposet **Bim**, and can be used as a way of axiomatizing potential functionality in general biposets. But they are also

the defining conditions for internal adjoint pairs.

Two opposed terms $y \xrightarrow{r} x$ and $y \xleftarrow{s} x$ form an *adjoint pair* of terms or an *adjunction*, denoted by $y \xrightarrow{r \dashv s} x$, when they satisfy the "unit inequality" $y \preceq s \circ r$ and the "counit inequality" $s \circ r \preceq x$. This axiomatizes "functionality" of **P**-terms. The term r is called the *left adjoint* and the term s is called the *right adjoint* in the adjunction $r \dashv s$. It is easy to show that right adjoints (and left adjoints) are unique, when they exist: if $y \xrightarrow{r \dashv s_1} x$ and $y \xrightarrow{r \dashv s_2} x$ then $s_1 = s_2$. Denote the unique right adjoint of $y \xrightarrow{r} x$ by $y \xleftarrow{r^{\mathrm{op}}} x$. A *functional* **P**-*term* is a **P**-term with a right adjoint. We usually use the notation $y \xrightarrow{f \dashv f^{\mathrm{op}}} x$ for functional terms. For any adjoint pair $y \xrightarrow{f \dashv f^{\mathrm{op}}} x$: when the unit is equality $y = f \circ f^{\mathrm{op}}$ they are a *coreflective pair*; when the counit is equality $f^{\mathrm{op}} \circ f = x$ they are a *reflective pair*; and when both unit and counit are equalities they are an *inverse pair*. For any functional term $y \xrightarrow{f \dashv f^{\mathrm{op}}} x$: the adjunction $f \dashv f^{\mathrm{op}}$ is a coreflection iff f is a monomorphism (iff f^{op} is an epimorphism); the adjunction is a reflection iff f is an epimorphism (iff f^{op} is a monomorphism); and the adjunction is an inversion iff f is an isomorphism (iff f^{op} is an isomorphism), iff $f^{\mathrm{op}} = f^{-1}$ is the two-sided inverse of f. Again we see that (in this case, internal) adjointness is a kind of generalized inverse. An internal coreflective pair $y \xrightarrow{i \dashv p} x$ is also called a *subtype* of x. Adjoint pairs compose in the obvious way: $(g \dashv g^{\mathrm{op}}) \circ (f \dashv f^{\mathrm{op}}) = (g \circ f) \dashv (f^{\mathrm{op}} \circ g^{\mathrm{op}})$, and $(x \dashv x)$ is the identity adjoint pair at x. So each biposet **P** has an associated adjoint pair category $\mathbf{P}^{\dashv}$, whose objects are **P**-types and whose arrows are **P**-adjunctions. Equivalently, $\mathbf{P}^{\dashv}$-arrows are just functional **P**-terms. There is an inclusion functor $\mathbf{P}^{\dashv} \xrightarrow{\mathrm{Inc}} \mathbf{P}$. The construction $(\ )^{\dashv}$ can be described as either "internal dialecticality" or "functionality".

In objective dialectics, since dialectical contradictions are represented by adjunctions, systems of dialectical contradictions are represented by diagrams in (pseudofunctors into) the category **Adj** whose objects are small categories and whose morphisms are adjoint pairs of functors. We call such a (pseudo)functor $\mathbf{P} \xrightarrow{E} \mathbf{Adj}$ a *dialectical base* or an *indexed adjointness*, and use the notation $E(y \xrightarrow{r} x) = (E^r \dashv E_r) \colon E(y) \to E(x)$. A dialectical base can be split into its *direct flow aspect* $\mathbf{P} \xrightarrow{E(\ )} \mathbf{Cat}$ and its *inverse flow aspect* $\mathbf{P}^{\mathrm{op}} \xrightarrow{E(\ )} \mathbf{Cat}$. Objects of **P** are called *types* and arrows of **P** are called *terms*. A *dialectical system* $y \xrightarrow{\iota, o} x$ is a graph in **P**, with inverse flow specifier ι and direct flow specifier o. Dialectical systems are the "motors of nature" specifying the dialectical motion of structured entities, and a dialectical base provides the "motive power" for this motion (from a dialectical point-of-view "motion" is synonymous with "transformation"). In this paper we are chiefly concerned with dialectical bases of preorders. Here a dialectical base

$\mathbf{P} \xrightarrow{E} \mathbf{adj}$ factors through the category **adj** of preorders and adjoint pairs of monotonic functions, and direct flow $\mathbf{P} \xrightarrow{E(\ )} \mathbf{PO}$ and inverse flow $\mathbf{P}^{\mathrm{op}} \xrightarrow{E(\ )} \mathbf{PO}$ map to preorders (and usually semilattices). Any functional term $y \xrightarrow{f} x$ in a biposet $\mathbf{P}$ defines a *direct image* monotonic function $\mathbf{P}[y,y] \xrightarrow{\mathbf{P}f} \mathbf{P}[x,x]$ defined by $\mathbf{P}f(q) \stackrel{\mathrm{df}}{=} f^{\mathrm{op}} \circ q \circ f$ for endoterms $y \xrightarrow{q} y$, and an *inverse image* monotonic function $\mathbf{P}[y,y] \xleftarrow{\mathbf{P}_f} \mathbf{P}[x,x]$ defined by $\mathbf{P}_f(p) \stackrel{\mathrm{df}}{=} f \circ p \circ f^{\mathrm{op}}$ for endoterms $x \xrightarrow{p} x$. It is easy to check that direct and inverse image form an adjoint pair of monotonic functions $\mathbf{P}(y \xrightarrow{f} x) = \mathbf{P}[y,y] \xrightarrow{\mathbf{P}f \dashv \mathbf{P}_f} \mathbf{P}[x,x]$ for each functional $\mathbf{P}$-term $y \xrightarrow{f \dashv f^{\mathrm{op}}} x$. The construction $\mathbf{P}$, mapping types to their poset of endoterms $\mathbf{P}(x) = \mathbf{P}[x,x]$ and mapping functional $\mathbf{P}$-terms to their adjoint pair of direct/inverse image adjunction, is a dialectical base (indexed adjointness) $\mathbf{P}^{\dashv} \xrightarrow{\mathbf{P}} \mathbf{adj}$.

Bisemilattices. The structural aspect of both the intuitionistic and classical semantics of dialectical logic is defined in terms of bisemilattices. A *join bisemilattice* or *semiexact biposet* is a biposet whose homsets are finitely complete (join-)semilattices and whose composition is finitely (join-)continuous. Horizontally the term "semilattice-valued category" might be indicated, but vertically from a bicategorical viewpoint the term "bisemilattice" seems appropriate. In more detail, a join bisemilattice $\mathbf{P} = \langle \langle \mathbf{P}, \preceq, \circ, \mathrm{Id} \rangle, \vee, \bot \rangle$ consists of the data and axioms of a biposet $\mathbf{P} = \langle \mathbf{P}, \preceq, \circ, \mathrm{Id} \rangle$, plus the following. For any two parallel terms $y \xrightarrow{s,r} x$ there is a *join* term $y \xrightarrow{s \vee r} x$ satisfying $s \vee r \preceq \preceq_{y,x} t$ iff $s \preceq_{y,x} t$ and $r \preceq_{y,x} t$. For any pair of types y and x there is an *empty* (or *bottom*) term $y \xrightarrow{\bot_{y,x}} x$ satisfying $\bot_{y,x} \preceq r$. The tensor product is finitely (join-) continuous (distributive w.r.t. finite joins) on the right and the left, $s \circ (r_1 \vee \cdots \vee r_n) = (s \circ r_1) \vee \cdots \vee (s \circ r_n)$ and $(s_1 \vee \cdots \vee s_m) \circ r = (s_1 \circ r) \vee \cdots \vee \vee (s_m \circ r)$ for any natural numbers n and m, including 0. A join bisemilattice with one object (universal type) is called a *monoidal join semilattice*. For any $\mathbf{P}$-term $y \xrightarrow{r} x$ the associated closed-below subset $\bot(r)$ of terms orthogonal to r is also closed under finite joins: $\bot_{x,y} \in \bot(r)$, and if $s_1, s_2 \in \bot(r)$ then $(s_1 \vee s_2) \in \bot(r)$ also. So $\bot(r)$ is an order ideal called the *orthogonality ideal* of r. If $\mathbf{P}$ is a join bisemilattice, then the opposite biposet $\mathbf{P}^{\mathrm{op}}$ is also a join bisemilattice. A *meet bisemilattice* is a biposet whose co-dual biposet is a join bisemilattice; that is, whose homsets are finitely complete (meet-) semilattices and whose composition is finitely (meet-)continuous. For any two parallel terms $y \xrightarrow{s,r} x$ there is a *meet* term $y \xrightarrow{s \wedge r} x$ satisfying $t \preceq_{y,x} s \wedge r$ iff $t \preceq_{y,x} s$ and $t \preceq_{y,x} r$. For any pair of types y and x there is a *full* (or *top*) term $y \xrightarrow{\top_{y,x}} x$ satisfying $r \preceq \top_{y,x}$. A *morphism of join bisemilattices* $\mathbf{P} \xrightarrow{H} \mathbf{Q}$ is a functor which preserves homset order and finite homset joins. A *bilattice*

or *exact biposet* is a join bisemilattice whose homsets are lattices. Note: a bilattice is not necessarily a meet bisemilattice.

To recapitulate, a join bisemilattice $\mathbf{P} = \langle \mathbf{P}, \preceq, \circ, \mathrm{Id}, \vee, \perp \rangle$ is the central structural notion in dialectical logic. It should be viewed as a direct generalization of a distributive lattice $L = \langle L, \leqq, \wedge, \top, \vee, \perp \rangle$. The generalization occurs in two different senses. (1) A join bisemilattice is a distributed structure: the notion of types is included, and the lattice operations are distributed over and between types. (2) The lattice meet $s \wedge r$ is replaced by the tensor product $s \circ r$, and the top (meet unit) $\top$ is replaced by the identities $x \xrightarrow{x} x$. Since a lattice meet is associative, unital, commutative, idempotent, and unit bounded, whereas a tensor product is only associative and unital, we see that commutativity, idempotency and unit-boundedness are discarded globally in the generalization. However, these three properties are incorporated in dialectical logic in two distinct ways. On the one hand, in the object aspect of dialectical logic the laws of idempotency and partiality (unit-boundedness) are incorporated locally in the idea of comonoid (see appendix). These local comonoidal contexts are standard contexts. Comonoidal structures define the generalized topological notions of interior and closure of terms, which are the modalities of affirmation and consideration from linear logic [Girard]. In axiomatics and proof theory, the idempotency and partiality axioms are known as contraction and weakening. On the other hand, in the construction of the classical context from the intuitionistic context, a natural weakened form of commutativity, called quasisymmetry, is found to be essential. Moreover, in the object aspect of classical dialectical logic, quasisymmetry is equivalent to internal (topological) dialecticality!

A *complete Heyting category* or *complete bilattice*, abbreviated cHc, is the same as a complete join bisemilattice; that is, a join bisemilattice $\mathbf{H}$ whose homsets are complete join semilattices (arbitrary joins exist) and whose tensor product is join continuous (completely distributive w.r.t. joins) on the right and the left, $s \circ (\vee_i r_i) = \vee_i (s \circ r_i)$ and $(\vee_j s_j) \circ r = \vee_j (s_j \circ r)$. Since the homset $\mathbf{H}[x, z]$ is a complete lattice and the left tensor product $\mathbf{H}[x, z] \xrightarrow{r\circ} \mathbf{H}[y, z]$ is continuous, it has (and determines) a right adjoint $\mathbf{H}[x, z] \xleftarrow{r\backslash} \mathbf{H}[y, z]$ called *left tensor implication*, and defined by $r \backslash t \stackrel{\mathrm{df}}{=} \bigvee \{x \xrightarrow{s} z \mid r \circ s \preceq_{y,z} t\}$. Adjointness means that left tensor product and left tensor implication satisfy the dialectical axiom $r \circ s \preceq_{y,z} t$ iff $s \preceq_{x,z} r \backslash t$. Similarly, the right tensor product $\mathbf{H}[z, y] \xrightarrow{\circ r} \mathbf{H}[z, x]$ has (and determines) a right adjoint $\mathbf{H}[z, y] \xleftarrow{t/r} \mathbf{H}[z, x]$ called *right tensor implication*, and defined by $s / r \stackrel{\mathrm{df}}{=} \bigvee \{z \xrightarrow{t} y \mid t \circ r \preceq_{z,x} s\}$. Adjointness means that right tensor product and right tensor implication satisfy the dialectical axiom $t \circ r \preceq_{z,x} s$ iff $t \preceq_{z,y} s / r$. A complete Heyting category with one object (universal type) is called a *complete Heyting monoid* [Birkhoff, Henkin] $\mathbf{M} = \langle M, \preceq, \circ, e, \backslash, /, \vee \vee, \perp, \wedge, \top \rangle$. If $\mathbf{M}$ is symmetric, then the two tensor implications are one:

$\Rightarrow \stackrel{\text{df}}{=} \backslash - = /-$. A complete symmetric Heyting monoid is known as a *closed* (*monoidal*) *poset*.

Examples. Complete Heyting categories are everywhere. The datatype $\mathbf{2} = \langle \{0,1\}, \leqq, \wedge, 1, \Rightarrow, \vee, 0 \rangle = \mathcal{P}(\mathbf{1})$ of Boolean values is a complete Heyting monoid, whose tensor product is the homset lattice meet $\wedge =$ **and** with unit $1 =$ **true**, and whose homset Boolean sum is $\vee =$ **or** with bottom $\bot = 0 = $ $=$ **false**. The powerset datatype $\mathcal{P}(A) = \langle \mathcal{P}(A), \subseteq, \cap, A, \Rightarrow, \cup, \emptyset \rangle$ of subsets of a fixed set A is a complete Heyting monoid. More generally, any complete Heyting algebra $\mathbf{M} = \langle M, \preceq, \wedge, \top, \Rightarrow, \vee, \bot \rangle$ is the same as a complete *Cartesian* Heyting monoid, where tensor product coincides with homset lattice meet $s \circ r = s \wedge r$. The category **Rel** is a complete Heyting category. Given a monoid $\mathbf{M} = \langle M, \circ, e, \rangle.$, the category of formal $\mathbf{M}$-languages $\mathcal{P}(\mathbf{M})$ is a complete Heyting monoid, where tensor product is language concatenation $L \bullet K$ with unit $\{e\}$, and the two tensor implications are (left and right) language division or cut $L \setminus K \stackrel{\text{df}}{=} \{m \in M \mid \forall_{n \in M} \text{ if } n \in L \text{ then } n \circ m \in K\}$. In particular, given an alphabet A, the category of formal A-languages $\mathcal{P}(A^*)$ is a complete Heyting monoid (the free complete Heyting monoid over the set A). The extended nonnegative real numbers $\mathbf{R} = \langle [0,\infty], \geqq, +, 0, -, \wedge, \infty, \vee, 0 \rangle$ with opposite order is a complete (non-Cartesian) Heyting monoid, where tensor product is numerical sum $s + r$ with unit 0, and tensor implication is numerical difference $s - r \stackrel{\text{df}}{=} s - r$ if $s \geqq r$, $= 0$ otherwise. There is a complete Heyting monoid $\mathcal{P}(\mathbf{R})$ associated with the extended nonnegative real numbers $\mathbf{R}$, whose morphisms $0 \stackrel{R}{\to} 0$ are subsets of reals $R \subseteqq [0,\infty]$ with $\bot_{0,0} = \emptyset$ and $\top_{0,0} = [0,\infty]$, whose homset order is the closed-above order $S \preceq$ $\preceq R$ when $S \subseteqq \uparrow(R)$, whose composition is defined pointwise by $S \circ R \stackrel{\text{df}}{=} \{s +$ $+ r \mid s \in S, r \in R\}$, and whose identity is $0 \stackrel{\{0\}}{\to} 0$. The singleton operator $\mathbf{R} \stackrel{\{\,\}}{\longrightarrow} \mathcal{P}(\mathbf{R})$ functorially embeds $\mathbf{R}$ into $\mathcal{P}(\mathbf{R})$. The infimum operator $\wedge$ is a functor $\mathcal{P}(\mathbf{R}) \stackrel{\wedge}{\longrightarrow} \mathbf{R}$, and (on the single homset) infimum reflects $\wedge \dashv \{\,\}$ the powerset of reals $\mathcal{P}(\mathbf{R})$ into the reals $\mathbf{R}$. The examples $\mathcal{P}(A^*)$ and $\mathcal{P}(\mathbf{R})$ motivate and are special cases of the following important construction. Just as every set C has an associated subset Heyting algebra $\mathcal{P}(C)$, so also every category $\mathbf{C}$ has an associated *subset category* $\mathcal{P}(\mathbf{C})$, whose objects are $\mathbf{C}$-objects, and whose arrows are subsets of homsets: $y \stackrel{R}{\to} x$ when $R \subseteqq \mathbf{C}[y,x]$. So $\mathcal{P}(\mathbf{C})[y,x] = \mathcal{P}(\mathbf{C}[y,x])$ with $\bot_{y,x} = \emptyset$ and $\top_{y,x} = \mathbf{C}[y,x]$. The tensor product in $\mathcal{P}(\mathbf{C})$ is defined pointwise, $S \circ R \stackrel{\text{df}}{=} \{z \stackrel{s \cdot_C r}{\longrightarrow} x \mid s \in S, r \in R\}$, generalizing the concatenation of formal languages and the addition of nondeterministic reals. The identity at x is the singleton set $x \stackrel{\{x\}}{\longrightarrow} x$, which can be identified with x itself. The left tensor implication is defined by $R \backslash - T \stackrel{\text{df}}{=} \{x \stackrel{s}{\to} z \mid (\forall r)$ if $r \in R$ then $r \cdot_C s \in T\}$ for any two $\mathcal{P}(\mathbf{C})$-arrows $y \stackrel{R}{\to} x$ and $y \stackrel{T}{\to} z$, and the

right tensor implication is defined dually. The Booleans are the "simplest" subset category $\mathbf{2} = \mathcal{P}(\mathbf{1})$.

More generally, every biposet $\mathbf{P}$ has an associated *closure subset category* $\mathcal{P}(\mathbf{P})$, whose arrows, tensor product, and identities are as in the unordered (identity order) case, and whose homset order is the closed-below order $S \preceq R$ when $S \subseteq\!\!\!| \downarrow(R)$. The definition of the implications follow from the continuity of the tensor product: the left tensor implication is $R \backslash T \stackrel{\text{df}}{=} \bigcup\{x \xrightarrow{S} z \mid R \circ S \preceq \preceq T\}$, and the right tensor implication is defined dually. Since every category $\mathbf{C}$ is a biposet with the identity order on homsets, the subset construction $\mathcal{P}(\mathbf{C})$ is a special case of the closure subset construction. It is easiest and most natural to define closure subset categories. Furthermore, this accords exactly with the appropriate generalization when biposets (or better, bipreorders) are replaced by bicategories. However, it is standard practice to use partial orders and closed subsets of terms. Any closure subset category $\mathcal{P}(\mathbf{P})$ has an associated *closed subset category* $\mathcal{K}(\mathbf{P})$, whose objects are the principal ideals $\{\downarrow(x) \mid x \text{ a } \mathbf{P}\text{-type}\}$, whose arrows $\downarrow(y) \xrightarrow{R} \downarrow(x)$ are closed-below subsets of terms $R \subseteq\!\!\!| \mathbf{P}[y, x]$ and $R = \downarrow(R)$, whose homset order is subset inclusion $S \preceq R$ when $S \subseteq\!\!\!| R$, and whose tensor product is the closure of the $\mathcal{P}(\mathbf{P})$-composition $S \circ R \stackrel{\text{df}}{=} \downarrow(\{z \xrightarrow{s \circ r} x \mid s \in S, r \in R\})$. The definition of the implications is as above $R \backslash T \stackrel{\text{df}}{=} \bigcup\{\downarrow(x) \xrightarrow{S} \downarrow(z) \mid R \circ S \preceq T\}$. For any biposet $\mathbf{P}$, the closed subset category $\mathcal{K}(\mathbf{P})$ is a complete Heyting category. For any $\mathbf{P}$-term $y \xrightarrow{r} x$ the orthogonality ideal is a term $x \xrightarrow{\perp(r)} y$ in $\mathcal{K}(\mathbf{P})$. In fact, orthogonality is a contravariant lax functor, $\perp(x) = \downarrow x$ and $\perp(r) \circ \perp(s) \subseteq\!\!\!| \perp(s \circ r)$, which is also hom-set contravariant, if $s \preceq r$ then $\perp(r) \subseteq\!\!\!| \perp(s)$.

Type sums. The closure subset construction $\mathcal{P}(\mathbf{P})$ does not capture the notion of "relational structures" completely. Although it introduces nondeterminism on the arrows, it leaves the objects alone. Type sums introduce distributivity on objects in a constructive fashion. We give a brief survey of type sums here.

A popular "external" model for predicates in logic is provided by subtypes. These are often constructed by a factorization/inclusion adjointness on slice categories of functional terms. Subtypes are closely connected with the "internal" model for predicates called comonoids (discussed in the appendix). For any type x, an *x-subtype* $y \xrightarrow{i \dashv p} x$ is another name for an internal coreflective pair $i \dashv p$ between y and x; that is, $y = i \circ p$ and $p \circ i \preceq x$. The interior term $x \xrightarrow{p \circ i} x$ is the comonoid associated with the subtype. We can define the usual subtype order between any two x-subtypes $y \xrightarrow{i \dashv p} x$ and $z \xrightarrow{j \dashv q} x$ as $\langle y, i \rangle \preceq \langle z, j \rangle$ when there exists a functional term $y \xrightarrow{h \dashv h^{\text{op}}} z$ such that $i = h \circ j$ and $q \circ h^{\text{op}} = p$. The largest x-subtype is the identity $x \xrightarrow{x \dashv x} x$. A term $z \xrightarrow{s} y$ is an (external) source *subterm* of a term $y \xrightarrow{r} x$, when $s = i \circ r$ for some

source subtype $z \xrightarrow{i \dashv p} y$. Two terms $z \xrightarrow{s} x$ and $y \xrightarrow{r} x$ with common target type x satisfy the *domain(-of-definition) order* $s \sqsubseteq r$ when z is a subtype of y mediated by the coreflective pair $z \xrightarrow{i \dashv p} y$ and $s \preceq i \circ r$. A more complete axiomatization of subtypes and comonoids is given in [Kent89].

The *empty type* 0 is a special type such that for any type x there are unique terms between x and 0 in either direction. So 0 is an initial type, satisfying the condition $0 \xrightarrow{r} x$ implies $r = \bot_{0,x}$; and 0 is a terminal type, satisfying the condition $x \xrightarrow{r} 0$ implies $r = \bot_{x,0}$. A type that is both initial and terminal is a null type. The null type 0 is the "empty sum", the sum of the empty collection of types. For any pair of types y and x, the bottom term $y \xrightarrow{\bot_{y,x}} x$ is the composition $\bot_{y,x} = \bot_{y,0} \circ \bot_{0,x}$. The empty type $0 \xrightarrow{\bot_{0,x} \dashv \bot_{x,0}} x$ is the smallest subtype of any type x, and its associated comonoid is the smallest comonoid. Given two types y and x, the *sum* of y and x is a composite type $y \oplus x$ having y and x as disjoint subtypes $y \xrightarrow{i_y \dashv p_y} y \oplus x \xleftarrow{i_x \dashv p_x} x$ which cover $y \oplus x$. So $y \oplus x$ comes equipped with two *injection terms* $y \xrightarrow{i_y} y \oplus x \xleftarrow{i_x} x$ and two *projection terms* $y \xleftarrow{p_y} y \oplus x \xrightarrow{p_x} x$ which satisfy the "comonoid covering equation" $(p_y \circ i_y) \vee (p_x \circ i_x) = y \oplus x$ stating that the subtype comonoids cover the sum type, and satisfy the "subtype disjointness equations" $i_y \circ p_y = y$, $i_y \circ p_x = \bot_{y,x}$, $i_x \circ p_y = \bot_{x,y}$, and $i_x \circ p_x = x$, or the "comonoid disjointness equation" $(p_y \circ i_y) \wedge (p_x \circ i_x) = \bot_{y \oplus x}$ stating that the subtype comonoids partition the sum type.

Equivalently, the sum type $y \oplus x$ is both a coproduct via the injections and a product via the projections of the types y and x. Given any pair of terms $y \xrightarrow{t} z \xleftarrow{s} x$ there is a unique term $y \oplus x \xrightarrow{[t,s]} z$, called the sum *source pairing* of t and s, which satisfies the source pairing conditions $i_y \circ [t,s] = t$ and $i_x \circ [t,s] = s$. Just define $[t,s] \stackrel{\text{df}}{=} (p_y \circ t) \vee (p_x \circ s)$. These properties say that the sum $y \oplus x$ is a coproduct. Equivalently, any term $y \oplus x \xrightarrow{r} z$ satisfies the "subterm covering condition" $r_y \vee r_x = r$ and the "subterm disjointness condition" $r_y \wedge r_x = \bot_{y \oplus x, z}$, where the y-th and x-th internal source subterms of r are defined by $r_y \stackrel{\text{df}}{=} (p_y \circ i_y) \circ r$ and $r_x \stackrel{\text{df}}{=} (p_x \circ i_x) \circ r$. Dually, given any pair of terms $y \xleftarrow{t} z \xrightarrow{s} x$ there is a unique term $z \xrightarrow{\langle t,s \rangle} y \oplus x$, called the sum *target pairing* of t and s, which satisfies the target pairing conditions $\langle t,s \rangle \circ p_y = t$ and $\langle t,s \rangle \circ p_x = s$. Just define $\langle t,s \rangle \stackrel{\text{df}}{=} (t \circ i_y) \vee (s \circ i_x)$. These properties say that the sum $y \oplus x$ is a product. Equivalently, any term $z \xrightarrow{r} y \oplus x$ satisfies the "subterm covering condition" $r^y \vee r^x = r$ and the "subterm disjointness condition" $r^y \wedge r^x = \bot_{y \oplus x, z}$, where the y-th and x-th internal target subterms of r are defined by $r^y \stackrel{\text{df}}{=} r \circ (p_y \circ i_y)$ and $r^x \stackrel{\text{df}}{=} r \circ (p_x \circ i_x)$. An object which is both a product and a coproduct of two other objects is called a *biproduct*. So type sums are biproducts. A join bisemilattice **P** is said to have *type sums*

or *biproducts* when type sums exist for any (finite) collection of types.

Domains/totality. The "action" of a term $y \xrightarrow{r} x$ is concentrated in and localized to a "locus of activity", a source subtype called the domain-of-definition of r (and a target subtype called the range of r). This domain is a kind of "effect" or "read-out" of a term r, and defines predicate transformation [Kent89] so that r becomes a predicate transformer. There are two approaches for formulating this.

One approach regards the notion of total term as fundamental, and domain-of-definition as derived. In this approach a term $y \xrightarrow{r} x$ is defined to be *total* when $s \circ r = \bot_{z,x}$ implies $s = \bot_{z,y}$ for any term $z \xrightarrow{s} y$. We then axiomatize the notion of domain-of-definition by assuming that inclusion of total terms has a right adjoint right inverse $(\ )^\dagger$ called the *totalization* or *total subterm operator* at x, forming a coreflective pair $\mathrm{Inc} \dashv (\ )^\dagger$ with $\mathrm{Inc} \cdot (\ )^\dagger = \mathrm{Id}$. This means that $t \sqsubseteq r$ iff $t \sqsubseteq r^\dagger$ for any total term $z \xrightarrow{t} x$ and any term $y \xrightarrow{r} x$; moreover, $t^\dagger = t$ for any total term t. Equivalently, $r^\dagger$ is the largest total term under r in the domain order: (1) $r^\dagger \sqsubseteq r$ and (2) $t \sqsubseteq r$ implies $t \sqsubseteq r^\dagger$ for total t. So, there is a y-subtype $d \xrightarrow{i \dashv p} y$ called the *domain subtype* of r, such that $r^\dagger \preceq i \circ r$. Since total terms are closed above we must have equality $r^\dagger = i \circ r$. The associated r-subterm $r^\dagger$ is called the *totalization* of r. The domain subtype $d \xrightarrow{i \dashv p} y$ is the y-subtype where the term $y \xrightarrow{r} x$ has non-nil action. It is the largest y-subtype whose associated r-subterm is total, in the sense that any other such subtype factors through the domain subtype. We need additional axioms to ensure that any term r is recoverable from its totalization by the identity $r = p \circ r^\dagger$.

Another, perhaps better, approach regards the notion of domain-of-definition as fundamental, and defines totalness as a derived notion. The *domain subtype* of any term $y \xrightarrow{r} x$ is the source subtype $\partial_0(r) = d_r \xrightarrow{i_r \dashv p_r} y$ which satisfies the axioms: (1) "minimality" $z \succeq \partial_0(r)$ iff $p \circ i \circ r = r$ for any source subtype $z \xrightarrow{i \dashv p} y$; (2) "composition" $\partial_0(s \circ r) = \partial_0(s \circ p_r)$ for any composable term $z \xrightarrow{s} y$; and (3) "monotonicity" $r \preceq r'$ implies $\partial_0(r) \preceq \partial_0(r')$ for any parallel term $y \xrightarrow{r'} x$. Define the *totalization* of r to be the r-subterm $r^\dagger \stackrel{\mathrm{df}}{=} i_r \circ r$. A term $y \xrightarrow{r} x$ is *total* when its domain is the largest source subtype, the entire source type $\partial_0(r) = y$. Some identities for the domain operator ∂_0 are: types are their own domain $\partial_0(x) = x$; the totalization is total, since $\partial_0(r^\dagger) = \partial_0(i_r \circ r) = \partial_0(i_r \circ p_r) = \partial_0(d_r) = d_r$; functional terms $y \xrightarrow{f \dashv f^{\mathrm{op}}} x$ are total, since the counit inequality $y \preceq f \circ f^{\mathrm{op}}$ implies $y = \partial_0(y) \preceq \partial_0(f \circ f^{\mathrm{op}}) = \partial_0(f \circ p_{f^{\mathrm{op}}}) \preceq \partial_0(f \circ x) = \partial_0(f) \preceq y$; in particular, subtypes are total $\partial_0(y \xrightarrow{i \dashv p} x) = y$; domain subtypes are their own domain, since $\partial_0(p_r) = \partial_0(p_r \circ d_r) = \partial_0(p_r \circ r^\dagger) = \partial_0(r) = d_r$; only zero has empty domain $\partial_0(r) = 0 \xrightarrow{\bot_{0,y} \dashv \bot_{y,0}} y$ iff $r = 0_{y,x}$

for any term $y \xrightarrow{r} x$; and given any two total terms $z \xrightarrow{s} y$ and $y \xrightarrow{r} x$, the composite term $z \xrightarrow{s \circ r} x$ is also total, since $\partial_0(s \circ r) = \partial_0(s \circ p_r) = \partial_0(s \circ y) = \partial_0(s) = z$.

Total terms are close above w.r.t. term entailment order. Since functional terms (in particular, identity terms) are total, and the composite of total terms are also total, total terms form a biposet $\mathbf{P}^\dagger$, a subbiposet of $\mathbf{P}$, $\mathbf{P}^\dashv \subseteqq \mathbf{P}^\dagger \subseteqq \mathbf{P}$, which is the homset order closure of $\mathbf{P}^\dashv$. So $\mathbf{P}^\dagger$ is a subbiposet $\mathbf{P}$, which preserves homset joins but usually does not have a bottom. Total terms in Heyting categories have been suggested [Hoare87] (although not by that name) as good models for programs (brief discussion in the section on Heyting categories).

Matrices and distributors. There is a cHc with type sums $\mathcal{M}(\mathbf{R})$ associated with the complete Heyting monoid of nonnegative reals $\mathbf{R} = \langle [0,\infty], \geqq, +, 0, -, \wedge, \infty, \vee, 0 \rangle$; whose objects are sets $X, Y, Z, \cdots$, whose morphisms $Y \xrightarrow{\phi} X$ are $Y \times X$-indexed collections of reals $\phi = \{\phi_{yx} \mid y \in Y, x \in X\}$ (that is, real-valued characteristic functions $Y \times X \xrightarrow{\phi} [0,\infty]$), whose composition $Z \xrightarrow{\psi \circ \phi} X$ for morphisms $Z \xrightarrow{\psi} Y$ and $Y \xrightarrow{\phi} X$ is $(\psi \circ \phi)_{zx} \stackrel{\text{df}}{=} \bigwedge_{y \in Y} [\psi_{zy} + \phi_{yx}]$, and whose identity $X \xrightarrow{X} X$ at X is defined by $X_{x'x} = 0$ if $x' = x$, $= \infty$ otherwise. Terms $Y \xrightarrow{\phi} X$ can be viewed as *fuzzy relations*, where ϕ_{yx} measures the degree of membership in ϕ, with $\phi_{yx} = 0$ asserting full (crisp) membership $(y,x) \in \phi$ and $\phi_{yx} = \infty$ asserting full nonmembership $(y,x) \notin \phi$. More generally, every cHc $\mathbf{H}$ has an associated *matrix category* $\mathcal{M}(\mathbf{H})$, whose objects are $\mathbf{H}$-*vectors* $\mathcal{X} = \langle X, |\ |_{\mathcal{X}} \rangle$ where X is an indexing (node) set and $X \xrightarrow{|\ |_{\mathcal{X}}} \mathrm{Obj}(\mathbf{H})$ is a (typing) function, whose arrows $\mathcal{Y} \xrightarrow{R} \mathcal{X}$ are $\mathbf{H}$-*matrices* where R is a $Y \times X$-indexed collection of $\mathbf{H}$-terms $R = \left(|y|_{\mathcal{Y}} \xrightarrow{r_{yx}} |x|_{\mathcal{X}} \mid \mid y \in Y, x \in X \right)$ (in other words, a generalized $\mathrm{Ar}(\mathbf{H})$-valued characteristic functions $Y \times X \xrightarrow{r} \mathrm{Ar}(\mathbf{H})$ compatible with source and target), whose homset order is pointwise order $(s_{yx}) \preceq (r_{yx})$ when $s_{yx} \preceq r_{yx}$ for all $y \in Y$ and $x \in X$, whose composition is matrix tensor product $(S \circ R)_{zx} = S_{zY} \circ R_{Yx} = \bigvee_{y \in Y} (s_{zy} \circ r_{yx})$ *"matrix tensor product"* for composable matrices $\mathcal{Z} \xrightarrow{S} \mathcal{Y}$ and $\mathcal{Y} \xrightarrow{R} \mathcal{X}$, whose identity at $\mathcal{X}$ is the diagonal matrix $\mathcal{X} \xrightarrow{\mathcal{X}} \mathcal{X}$ defined as identity $\mathbf{H}$-terms $\mathcal{X}_{xx} = |x|_{\mathcal{X}} \xrightarrow{|x|_{\mathcal{X}}} |x|_{\mathcal{X}}$ on the diagonal and zero (bottom) $\mathbf{H}$-terms $\mathcal{X}_{x'x} = |x'|_{\mathcal{X}} \xrightarrow{\perp} |x|_{\mathcal{X}}$ off the diagonal, and whose matrix tensor implications are $(S \backslash\!\!- R)_{zy} = S_{zX} \backslash\!\!- R_{yX} = \bigwedge_{x \in X} (s_{zx} \backslash\!\!- r_{yx})$ *"right matrix tensor implication"* and $(R -\!\!/\!\!- T)_{xz} = R_{Yx} -\!\!/\!\!- T_{Yz} = \bigwedge_{y \in Y} (r_{yx} -\!\!/\!\!- t_{yz})$ *"left matrix tensor implication"*. Matrices $Y \xrightarrow{R} X$ can be viewed as *fuzzy* $\mathbf{H}$-*relations*. For any cHc $\mathbf{H}$, the matrix category $\mathcal{M}(\mathbf{H})$ is a complete Heyting category for which

biproducts (type sums) exist. For the complete Cartesian Heyting monoid of Boolean values $\mathbf{2} = \langle \{0,1\}, \leqq, \wedge, 1, \Rightarrow, \vee, 0 \rangle = \mathcal{P}(\mathbf{1})$ the associated cHc with biproducts is $\mathcal{M}(\mathbf{2}) = \mathcal{M}(\mathcal{P}(\mathbf{1})) = \mathbf{Rel}$ the category of ordinary relations.

Every category $\mathbf{C}$ has an associated *distributor category* $\mathcal{D}(\mathbf{C})$ defined by $\mathcal{D}(\mathbf{C}) \stackrel{\text{df}}{=} \mathcal{M}(\mathcal{P}(\mathbf{C}))$. In more detail, $\mathcal{D}(\mathbf{C})$ is the category, whose objects are *distributed* $\mathbf{C}$*-objects* or $\mathbf{C}$*-vectors* $\mathcal{X} = \langle X, |\ |_{\mathcal{X}} \rangle$ as above, whose arrows $\mathcal{Y} \stackrel{R}{\rightarrow} \mathcal{X}$ are *distributed* $\mathbf{C}$*-arrows* or $\mathbf{C}$*-distributors* where $R \subseteqq Y \times \mathrm{Ar}(\mathbf{C}) \times X$ is a digraph between the underlying node sets consisting of compatible triples: if $(y, r, x) \in R$ then $|y|_{\mathcal{Y}} \stackrel{r}{\rightarrow} |x|_{\mathcal{X}}$ is a $\mathbf{C}$-arrow, whose tensor product is defined pointwise as $(S \circ R)_{z,x} \stackrel{\text{df}}{=} \bigcup_{y \in Y} [S_{zy} \circ R_{yx}]$, and whose identity at $\mathcal{X}$ is the $\mathbf{C}$-distributor $\mathcal{X} \stackrel{\text{df}}{=} \{(x, |x|_{\mathcal{X}}, x) \mid x \in X\} \subseteqq X \times \mathrm{Ar}(\mathbf{C}) \times X$ consisting (on the diagonal) of all the $\mathbf{C}$-identities indexed by $\mathcal{X}$. The (y,x)-th fiber of a $\mathcal{D}(\mathbf{C})$-term $\mathcal{Y} \stackrel{R}{\rightarrow} \mathcal{X}$, defined by $R_{yx} \stackrel{\text{df}}{=} \{y \stackrel{r}{\rightarrow} x \mid r \in R\}$, is a $\mathcal{P}(\mathbf{C})$-term $y \stackrel{R_{yx}}{\rightarrow} x$, and R is the disjoint union $R = \coprod_{y \in \mathcal{Y}, x \in \mathcal{X}} R_{yx}$ of its $\mathcal{P}(\mathbf{C})$-term fibers. For any category $\mathbf{C}$, the distributor category $\mathcal{D}(\mathbf{C})$ is a complete Heyting category for which biproducts (type sums) exist. The category of relations is the "simplest" distributor category $\mathbf{Rel} = \mathcal{D}(\mathbf{1})$. Since any category $\mathbf{C}$ has a unique functor $\mathbf{C} \stackrel{!}{\rightarrow} \mathbf{1}$ to the one-arrow category, every distributor category has a functor (morphism of distributor categories) $\mathcal{D}(\mathbf{C}) \stackrel{\mathcal{D}(!)}{\longrightarrow} \mathbf{Rel} = \mathcal{D}(\mathbf{1})$.

In distributor categories $\mathcal{D}(\mathbf{C})$ a comonoid W of type X is essentially a subobject (subset) $W \subseteqq X$, and so $\Omega(X) \cong \mathcal{P}(X)$. More generally, every biposet $\mathbf{P}$ has an associated *closure distributor category* $\mathcal{D}(\mathbf{P}) \stackrel{\text{df}}{=} \mathcal{M}(\mathcal{P}(\mathbf{P}))$, whose objects, arrows, tensor product and identities are as above, and whose homset order is the pointwise closed-below order. Given any set of attributes or sorts A, a signature $\Sigma = \{\Sigma_{y,a} \mid y \in \mathrm{multiset}(A), a \in A\}$ over A determines a term category $\mathbf{T}_\Sigma$, the initial algebraic theory over Σ, whose objects are multisubsets of A (arities, tuplings, etc.) and whose arrows are tuples of Σ-terms. A parallel pair of arrows $\mathcal{Y} \stackrel{S,R}{\rightarrow} \mathcal{X}$ in the distributor category $\mathcal{D}(\mathbf{T}_\Sigma^{\mathrm{op}})$ is a Horn clause logic program, whose predicate names are $\mathcal{X}$-nodes, whose clause names are $\mathcal{Y}$-nodes, whose clause-head atoms are (w.l.o.g.) collected together as S, whose clause-body atoms are collected together as R, and whose associated fixpoint operator (see appendix) is the inverse/direct flow composite $((\ )/R) \circ S$ defined on Herbrand interpretations with database scheme $\mathcal{X}$. In much of the logic of dialectical processes (in particular, for Girard's completeness theorem) closure subset categories suffice. However, for the constraint dialectic, the full nondeterminism and parallelism of distributor categories is essential.

2. Semantics

Flow is at the heart of computational and dynamic systems. From the calculi and semantics of processes comes the notion of process communication and process flow. From logic programming and Petri net theory comes the idea that flow is dialectical, in the sense of moving in both a direct and an inverse direction. Flow is the behavior of dialectical processes. Direct flow is modelled by a nonsymmetric tensor product, whereas inverse flow is modelled by both a left (reverse-time, source, quo-object) tensor implication and a right (forward-time, target, subobject) tensor implication (or tensor exponentiations). This bidirectional notion of flow is called the *flow* (or *motion*) *dialectic.*

Both dialectical logic and linear logic deal principally with the dynamical notions of *state* and *transitions* (involving "dialectically contradictory" activities [Kent87], such as the creation/destruction or production/consumption of values, often representing resources), whereas standard logic, both classical and intuitionistic, deals with the relatively static notion of monotonically increasing truth values (once true, true forever). Dialectical and linear logic are proper extensions of standard logic, relegating the Cartesianness of the standard fragment [Kent88] (weakening, contraction, etc.) to local contexts: that is, they have locally Cartesian-closed semantical structures. Presently linear logic requires the commutativity or symmetry of tensor product, in order to define a simpler semantics. However, the semantics of dialectical processes, which includes traditional process semantics, is not commutative. This argues strongly that commutativity should be excluded initially, and only included later when desired via a symmetrization construction on the nonsymmetric case. The semantics and logic of dialectical processes in this paper agrees with linear logic in subject studied and philosophy. They disagree in approach taken (I use a previously developed theory of dialectical systems) and in emphasis: linear logic emphasizes the importance of the linearity properties of implication and negation; whereas dialectical logic emphasizes the importance of the central dialectical contradiction (adjointness) between tensor product and tensor implication, thus giving logic a process interpretation. The logic of dialectical processes is more general than linear logic for two reasons: 1. dialectical logic is nonsymmetric (has a nonsymmetric tensor product operation) with linear logic a symmetric subcase; 2. linear logic is a typeless subcase of dialectical logic (all types are merged into one type).

Heyting categories. The full intuitionistic semantics of dialectical logic is defined in terms of Heyting categories. Concisely speaking, a *Heyting category* is a closed bilattice; that is, a bilattice **H** whose tensor product has right adjoints on both left and right. The underlying bilattice represents the structural aspect of a Heyting category, whereas the closedness property represents the dialectical or flow aspect.

In more detail, the flow aspect consists of the following data and axioms. For any two **H**-terms $y \xrightarrow{r} x$ and $z \xrightarrow{s} x$ with common target type there is a composite term $z \xrightarrow{s/r} y$ between their source types, defined by the dialectical axiom $t \circ r \preceq_{z,x} s$ iff $t \preceq_{z,y} s/r$, stating that the binary operation $/$ called *right tensor implication*, is right adjoint to tensor product on the right. Right tensor implication $/$, like all exponentiation or division operators including numerical ones is covariantly monotonic on the left and contravariantly monotonic on the right. This dialectical axiom, generalizing the deduction theorem of standard logic, defines the formal semantics of tensor implication $/$ in terms of tensor product $\circ$. From the dialectical axiom easily follows the inference rule of right modus ponens $(s/r) \circ r \preceq s$ and the inference rule $t \preceq (t \circ r)/r$. Also immediate from the axioms are the transitive, reflexive, mixed associative and unital laws: $(t/s) \circ (s/r) \preceq (t/r)$, $y \preceq (r/r)$, $t/(s \circ r) = (t/r)/s$, $(r/x) = r$. Right tensor implication allows each term $y \xrightarrow{r} x$ to specify a *right inverse flow* $\mathbf{H}[z,y] \xleftarrow{/r} \mathbf{H}[z,x]$ for each type z. The above mixed associative and unital laws say that right inverse flow $/r$ is (contravariantly) functorial in r with respect to the category **H**. Thus, each term r, using right tensor product and right tensor implication, specifies a "right dialectical base" for each type z. Dually, for any two **H**-terms $y \xrightarrow{r} x$ and $y \xrightarrow{t} z$ with common source type there is a composite term $x \xrightarrow{r\backslash t} z$ between their target types, defined by the dialectical axiom $r \circ s \preceq \preceq_{y,z} t$ iff $s \preceq_{x,z} r\backslash t$, stating that the binary operation $\backslash$ called *left tensor implication*, is right adjoint to tensor product on the left. Left tensor implication allows each term $y \xrightarrow{r} x$ to specify a *left inverse flow* $\mathbf{H}[x,z] \xleftarrow{r\backslash} \mathbf{H}[y,z]$ for each type z. The mixed associative and unital laws say that left inverse flow $r\backslash$ is (covariantly) functorial in r with respect to the category **H**, thus defining a "left dialectical base". Together the left and right implications satisfy the mixed associative law $s\backslash(t/r) = (s\backslash t)/r$. From both the left and right modus ponens, we get the derived rules $(r/r)\backslash r = r = r/(r\backslash r)$. Since tensor product is left adjoint on both left and right to tensor implication, it preserves arbitrary joins $s \circ (r \vee r') = (s \circ r) \vee (s \circ r')$, $s \circ \bot_{y,x} = \bot_{z,x}$, $(s \vee s') \circ r = (s \circ r) \vee (s' \circ r)$ and $\bot_{z,y} \circ r = \bot_{z,x}$. Since tensor implications are right adjoint to tensor product, they preserve arbitrary meets $r\backslash(t \wedge t') = (r\backslash t) \wedge (r\backslash t')$, $r\backslash \top_{y,z} = \top_{x,z}$, $(s \wedge s')/r = (s/r) \wedge (s'/r)$ and $\top_{z,x}/r = \top_{z,y}$. The two dialectical axioms assert that the bilattice **H** is closed.

For any functional Heyting term $y \xrightarrow{f \dashv f^{\mathrm{op}}} x$, tensor implication relates the adjoints by $f = f^{\mathrm{op}}\backslash x$ and $f^{\mathrm{op}} = x/f$. More generally, left f-product is equal to left f^{op}-implication $f \circ (\,) = f^{\mathrm{op}}\backslash(\,)$ and right f^{op}-product is equal to right f-implication $(\,) \circ f^{\mathrm{op}} = (\,)/f$, and we have the adjoint triples

$$f^{\mathrm{op}} \circ (\,) \dashv f \circ (\,) = f^{\mathrm{op}}\backslash(\,) \dashv f\backslash(\,)$$
$$(\,) \circ f \dashv (\,) \circ f^{\mathrm{op}} = (\,)/f \dashv (\,)/f^{\mathrm{op}}.$$

Such adjoint triples appear naturally in the dialectical view of dynamic logic called the standard aspect [Kent89], which discusses the equivalent notions of hyperdoctrines of comonoids and spannable dialectical flow categories. A Heyting category with one object (universal type) is called a *Heyting monoid* $\mathbf{M} = \langle M, \preceq, \circ, e, \backslash, /, \vee, \bot, \wedge, \top \rangle$. A preliminary version of Heyting monoid without homset lattice notions, was investigated early on [Lambek], and called *residuated preorder*. See also [Birkhoff, Henkin]. The opposite biposet $\mathbf{H}^{\mathrm{op}}$ is a Heyting category with implications switched. Since complete Heyting categories are Heyting categories, Heyting categories are ubiquitous; in particular, subset categories $\mathcal{P}(\mathbf{C})$ and distributor categories $\mathcal{D}(\mathbf{C})$ are Heyting categories.

Concurrent with the development of this paper, an algebraic theory for the "laws of programming" has been advocated [Hoare87], whose axioms are essentially those for Heyting categories; or more precisely, Heyting categories (in particular, cHc) with affirmation/consideration modalities and domain subtypes. The affirmation modality is defined in the appendix. The consideration modality is its order-theoretic dual. The topological notions of affirmation and consideration are discussed further in both the standard aspect and the object aspect of dialectical logic [Kent88, Kent89]. In the program interpretation, arbitrary Heyting terms represent program specifications, total Heyting terms represent programs, and either subtypes or comonoids (see appendix) represent conditions. Types represent local contexts for local states of the system. Term entailment order is interpreted as a measure of "nondeterminism" with $r \preceq s$ asserting that r is more deterministic than s. The top term $y \overset{\top_{y,x}}{\longrightarrow} x$ represents the worst (most nondeterministic) program, and functional terms represent fully deterministic (minimally nondeterministic) programs. The bottom term $y \overset{\bot_{y,x}}{\longrightarrow} x$, although deterministic, is not a program since its domain-of-definition is empty. The totalization $d \overset{r!}{\longrightarrow} x$ of a term $y \overset{r}{\longrightarrow} x$ is the least deterministic program (on the domain-of-definition) of that specification. In summary, the "Laws of Programming" can be interpreted in Heyting categories as follows (see p. 36).

More recently [Kent89] these laws (concerning structure and flow in Heyting categories) have been connected with the older program semantics which uses Hoare triples.

Tensor negation. Glivenko's theorem, defining the classical part of standard intuitionistic logic, seems to rely in part upon the symmetry (commutativity) of the Boolean product (lattice meet) in Heyting algebras. Recall that a biposet $\mathbf{P}$ is quasisymmetric when $r \perp s$ iff $s \perp_x r$ iff $r \perp_y s$ for all $\mathbf{P}$-types y and x and all opposed pairs of $\mathbf{P}$-terms $y \overset{r}{\longrightarrow} x$ and $y \overset{s}{\longleftarrow} x$. We can define quasisymmetry for $\mathbf{P}$-terms alone: a $\mathbf{P}$-term $y \overset{r}{\longrightarrow} x$ is *quasisymmetric* or *orthogonally balanced* when $s \perp_x r$ iff $r \perp_y s$ for all $\mathbf{P}$-terms $x \overset{s}{\longrightarrow} y$ opposed to r. I cannot overemphasize the importance of the notion of quasisymmetry,

"Laws of programming"		Heyting categories	
program specifications	S	terms	$y \xrightarrow{r} x$
programs	P	total terms	$y \xrightarrow{t} x$
conditions	b	comonoids	$u \in \Omega(x)$
		subtypes	$y \xrightarrow{i \dashv p} x$
nondeterminism order	$P \subseteq Q$	term entailment order	$r \preceq s$
sequential composition	$P;Q$	tensor product	$s \circ r$
nondeterministic choice	$P \bigcup Q$	Boolean sum	$s \vee r$
SKIP, the nop	II	identity (types-as-terms)	$x \xrightarrow{x} x$
ABORT, the worst program	$\bot$	top term	$y \xrightarrow{\top_{y,x}} x$
weakest prespecification	S/T	tensor implication	$t \backslash\!\!- s$
conditional or branch	$P \triangleleft b \triangleright Q$	derived expression	$(v \circ r) \vee (\sim v \circ s)$
	if b then P else Q	where $\sim v \stackrel{\text{df}}{=} (v \Rightarrow \bot_y) = (v \backslash\!\!- \bot_y)^\circ$ and $(\,)^\circ$ is the affirmation modality	
iteration or while-loop	$b \bullet P$	derived expression	$(u \circ r)^\bullet \circ \sim u$
	while b do P	where $(\,)^\bullet$ is the consideration modality	

especially in the object aspect of classical dialectical logic [Kent88]. Dually, a **P**-term $y \xrightarrow{r} x$ is *coquasisymmetric* when it is quasisymmetric in the codual $\mathbf{P}^{\mathrm{co}}$, which is **P** with the opposite homset order; that is, when $r \circ s \succeq_{y,y} y$ iff $s \circ r \succeq_{x,x} x$ for all **P**-terms $x \xrightarrow{s} y$ opposed to r. Identities are quasisymmetric, and quasisymmetric **P**-terms are closed under composition. The *center* of **P**, denoted by $\mathcal{Z}(\mathbf{P})$, is the sub-biposet consisting of all **P**-types and all quasisymmetric **P**-terms. All **P**-isomorphisms are quasisymmetric. Quasisymmetric **P**-terms are closed under arbitrary joins w.r.t. $\preceq$ (when they exist). When arbitrary joins of quasisymmetric terms exist, the center $\mathcal{Z}(\mathbf{P})$ is a kind of generalized topology with finite tensor products functioning as "finite intersections" and arbitrary Boolean sums (joins) functioning as "arbitrary unions" [Kent88]. For this reason quasisymmetric terms are also called $\mathcal{Z}(\mathbf{P})$-open terms.

Now let the biposet **P** be a Heyting category **H**. For any **H**-term $y \xrightarrow{r} x$, the *left x-dual* of r is $x \backslash\!\!- r$, the largest term with source x and target y which is semi-orthogonal to r at x: $(x \backslash\!\!- r) \bot_x r$, and if $s \bot_x r$ for $x \xrightarrow{s} y$ then $s \preceq_{x,y} x \backslash\!\!- r$. Dually, the *right y-dual* of r is $r \backslash\!\!- y$, the largest term with source x and target y which is semi-orthogonal to r at y. We have $r \bot s$ iff $(s \bot_x r$ and $r \bot_y s)$ iff $(s \preceq_{x,y} x \backslash\!\!- r$ and $s \preceq_{x,y} r \backslash\!\!- y)$ iff $s \preceq_{x,y} (r \backslash\!\!- y) \wedge (x \backslash\!\!- r)$. Define the *tensor negation* of the Heyting term $y \xrightarrow{r} x$ to be the term $\neg r = \neg_{yx} r \stackrel{\text{df}}{=} (r \backslash\!\!- y) \wedge (x \backslash\!\!- r)$. So for any Heyting term $y \xrightarrow{r} x$, the orthogonality ideal $\bot(r)$ is the principal ideal $\bot(r) = \downarrow(\neg r) = \downarrow((r \backslash\!\!- y) \wedge (x \backslash\!\!- r))$, and tensor negation $x \xrightarrow{\neg r} y$ is the largest (oppositely directed) term orthogonal to r: $\neg r = \top_{\bot(r)}$; or, phrased

as an equivalence, $r \perp s$ iff $s \preceq_{x,y} \neg r$. The definition of Boolean categories below uses this equivalence to axiomatize tensor negation without the need for tensor implications. The sense of this equivalence is that tensor negation is the "tensor complement" of r. So tensor negation in dialectical logic is entirely analogous to (and generalizes) Boolean negation in standard logic, where the Boolean negation of a Heyting element a is the largest element disjoint from a, $a \wedge b = 0$ iff $b \leqq \neg a$. Since tensor negation $\mathbf{H}[y,x] \xrightarrow{\neg_{yx}} \mathbf{H}[x,y]^{\mathrm{op}}$ is contravariantly monotonic, $s \preceq_{y,x} r$ implies $\neg r \preceq_{x,y} \neg s$, it is a dialectical negation. In more detail, since orthogonality is a symmetrical notion, $s \preceq_{x,y} \neg_{yx} r$ iff $r \perp s$ iff $r \preceq_{y,x} \neg_{xy} s$, tensor negation is a self-adjoint monotonic function $\neg_{yx} \dashv \neg_{xy}^{\mathrm{coop}}$. Since tensor negation $\neg$ is self-adjoint, it maps arbitrary joins to meets $\neg(\vee_i r_i) = \wedge_i(\neg r_i)$, which in the binary case gives the DeMorgan's law: $\neg(s \vee r) = \neg s \wedge \neg r$ and in the nullary case gives the law: $\neg \perp_{y,x} = \top_{x,y}$. We also have the derived rule $\neg_{zx}(s \circ r) = (r \backslash (z \backslash s)) \wedge ((x / r) / s)$. As remarked before, the generalized inverseness notion of an adjoint pair of terms $y \xrightarrow{r \dashv s} x$ forms a kind of polar-tension structure, since there is only one possible right adjoint $r \dashv s$ iff $s = r^{\mathrm{op}}$. However, the generalized inverseness notion of an orthogonal pair of terms $y \xrightarrow{r \perp s} x$ does not form a polar-tension structure. But we can make orthogonality that by assuming the existence of tensor negations: $y \xrightarrow{r \perp \neg r} x$ forms a kind of polar-tension structure, since there is only one possible tensor negation $r \perp s$ iff $s \preceq \neg r$. A subtype $y \xrightarrow{i \dashv p} x$ has only one kind of complement $\neg i = p = i^{\mathrm{op}}$ and $\neg p = \neg(i^{\mathrm{op}}) = i$, whereas a functional $\mathbf{H}$-term $y \xrightarrow{f} x$ has two kinds of complements: its tensor negation $x \xrightarrow{\neg f} y$ and its right adjoint $x \xrightarrow{f^{\mathrm{op}}} y$. In general, these two complements are related by $\neg f \preceq f^{\mathrm{op}} = x / f$ and $\neg(f^{\mathrm{op}}) \preceq f = f^{\mathrm{op}} \backslash x$. The two complements are identical $\neg f = f^{\mathrm{op}}$ iff $y \xrightarrow{f \dashv f^{\mathrm{op}}} x$ is a subtype.

A Heyting term $y \xrightarrow{r} x$ is quasisymmetric precisely when the left and right orthogonal duals coincide and equal the tensor negation $\neg r = x / r = r \backslash y$, since $s \circ r \preceq x$ iff $s \preceq x / r$ iff $s \preceq r \backslash y$ iff $r \circ s \preceq y$. For a quasisymmetric functional term $y \xrightarrow{f} x$, the two kinds of complements, tensor negation and right adjoint, are one: $\neg f = f^{\mathrm{op}}$ and $f = \neg f^{\mathrm{op}}$; so that, $y \xrightarrow{f \dashv f^{\mathrm{op}}} x$ is a subtype. This is an indication that quasisymmetry is a very strong and restrictive concept. This should be compared with the result in the object aspect of dialectical logic, that "quasisymmetry is equivalent to topological dialecticality". Tensor negation is contravariant lax functorial $\neg r \circ \neg s \preceq_{x,z} \neg(s \circ r)$, so that tensor negation and tensor product are related by the inequalities $s \circ r \preceq \preceq \neg\neg s \circ \neg\neg r \preceq \neg(\neg r \circ \neg s)$ and $s \circ r \preceq \neg\neg(s \circ r) \preceq \neg(\neg r \circ \neg s)$. A Heyting term $y \xrightarrow{r} x$ is *coquasisymmetric* when it is the tensor negation $r = \neg s$ of a quasisymmetric term $x \xrightarrow{s} y$. This notion of Heyting coquasisymmetry is close to, but not identical with, the notion of biposet coquasisymmetry above.

However, they agree on closed Heyting terms (see below). By definition tensor negation maps quasisymmetric terms into coquasisymmetric terms. A term $y \xrightarrow{r} x$ is an **H**-isomorphism iff its tensor negation is a categorical inverse: $\neg r \circ r = x$ and $r \circ \neg r = y$. Isomorphisms are both quasisymmetric and coquasisymmetric. For isomorphisms the tensor implications are expressible as $r \backslash t = \neg r \circ t$ and $s / r = s \circ \neg r$.

Double negation. Let **H** be a Heyting category. Let $\neg\neg$ symbolize double tensor negation, defined by $\neg\neg_{yx} r \stackrel{\text{df}}{=} \neg_{xy}(\neg_{yx} r)$ for any pair of types y and x, and any term $y \xrightarrow{r} x$. Double negation $\neg\neg$ is a local closure operator: "monotonic" $r \preceq_{y,x} s$ implies $\neg\neg r \preceq_{y,x} \neg\neg s$, "increasing" $r \preceq_{y,x} \neg\neg r$, and "idempotent" $\neg\neg(\neg\neg r) = \neg\neg r$. A term $y \xrightarrow{r} x$ is *double-negation closed* when $r = \neg\neg r$; or equivalently, when $r = \neg s$ for some term $x \xrightarrow{s} y$. Denote the collection of closed terms in $\mathbf{H}[y, x]$ by $\neg\neg\mathbf{H}[y, x]$. Then $\neg\neg\mathbf{H}[y, x]$ is a lattice, which is a meet-subsemilattice of the lattice $\mathbf{H}[y, x]$ with meets in $\neg\neg\mathbf{H}[y, x]$, called classical Boolean products, identical $\Delta_i r_i = \wedge_i r_i$ to meets in $\mathbf{H}[y, x]$, and joins in $\neg\neg\mathbf{H}[y, x]$, called classical Boolean sums defined (following Glivenko) as the double negation $\oplus_i r_i = \neg\neg(\vee_i r_i)$ of joins in $\mathbf{H}[y, x]$. Double negation $\mathbf{H}[y, x] \xrightarrow{\neg\neg} \neg\neg\mathbf{H}[y, x]$ reflects $\neg\neg \dashv \mathbf{Inc}$ arbitrary Heyting terms into closed terms. Identity terms (types) are closed, since $x = \neg x$. The smallest and largest closed terms from y to x are $0_{y,x} \stackrel{\text{df}}{=} \neg\neg\bot_{y,x} = \neg\top_{x,y}$ and $1_{y,x} \stackrel{\text{df}}{=} \neg\neg\top_{y,x} = \top_{y,x} = \neg\bot_{x,y} = \neg 0_{x,y}$, respectively. If **H** is a quasisymmetric category, then all functional terms are subtypes, all subtypes are double-negation closed, its functional part $\mathbf{H}^{\dashv}$ is a "preorderlike" category consisting only of subtype terms $y \xrightarrow{i \dashv p} x$, and the dialectical base $\mathbf{H}^{\dashv} \xrightarrow{\mathbf{H}} \mathbf{adj}$ is an "extension/restriction" base with direct image $\mathbf{H}[y, y] \xrightarrow{p \circ (\ ) \circ i} \mathbf{H}[x, x]$ being "extension to x" and inverse image $\mathbf{H}[y, y] \xleftarrow{i \circ (\ ) \circ p} \mathbf{H}[x, x]$ being "restriction to y". So, if we are interested in a general notion of "functionality" in Heyting categories (such as ordinary functions in **Rel** or functors in **Cat**), then we should not assume quasisymmetry.

If $y \xrightarrow{r} x$ is a quasisymmetric term, then $\neg\neg r = [y / (r \backslash y)] \wedge [(x / r) \backslash x]$ (in a quasisymmetric category $\neg\neg r = y / (r \backslash y) = (x / r) \backslash x$). If $y \xrightarrow{r} x$ is quasisymmetric, then $\neg\neg r$ is also quasisymmetric, since $p \circ \neg\neg r \preceq x$ implies $p \circ r \preceq x$ iff $p \preceq \neg r = \neg\neg\neg r$ implies $\neg\neg r \circ p \preceq y$.

LEMMA 1 (Functoriality). *Double negation is lax functorial on quasisymmetric terms:* $\neg\neg s \circ \neg\neg r \preceq_{z,x} \neg\neg(s \circ r)$ *for all composable pairs of quasisymmetric terms* $z \xrightarrow{s} y$ *and* $y \xrightarrow{r} x$.

PROOF. We prove something equivalent: for all composable pairs of quasisymmetric terms $z \xrightarrow{s} y$ and $y \xrightarrow{r} x$, $s \circ \neg\neg r \preceq_{z,x} \neg\neg(s \circ r)$ when s is double negation closed. By modus ponens on left and right $((x / r) / s) \circ s \circ \circ ((x / r) \backslash x) \preceq x$. So (1) $s \circ ((x / r) \backslash x) \preceq_{z,x} ((x / r) / s) \backslash x = (x / (s \circ r)) \backslash x$. On

the other hand $(y/(r\backslash y)) \circ (r\backslash\neg s) \circ (\neg s\backslash y) \preceq y$ by transitivity (used twice). But $s = \neg\neg s \preceq \neg s\backslash y$ since s is closed and quasisymmetric. So $(y/(r\backslash y)) \circ (r\backslash\neg s) \circ s \preceq y$. Again since s is quasisymmetric $s \circ (y/(r\backslash y)) \circ (r\backslash\neg s) \preceq z$. Hence, (2) $s \circ (y/(r\backslash y)) \preceq z/(r\backslash\neg s) = z/(r\backslash(s\backslash z)) = z/((s \circ r)\backslash z)$. Putting both facts together $s \circ \neg\neg r = s \circ [y/(r\backslash y)] \wedge [(x/r)\backslash x] \preceq [s \circ (y/(r\backslash y))] \wedge [s \circ (x/r)\backslash x)] \preceq [z/((s \circ r)\backslash z)] \wedge [(x/(s \circ r))\backslash x] = \neg\neg(s \circ r)$. Finally, $\neg\neg s \circ \neg\neg r \preceq \preceq_{z,x} \neg\neg(\neg\neg s \circ r) \preceq_{z,x} \neg\neg(\neg\neg(s \circ r)) = \neg\neg(s \circ r)$ by monotonicity and idempotency of $\neg\neg$. □

By rights this functoriality lemma should be called the "bottleneck lemma" since we need it [Girard] to prove associativity of the classical tensors defined below. The concept of quasisymmetry, although quite natural by itself, was motivated by this lemma.

Following Glivenko, in analogy with the definition of the classical Boolean connectives, the tensor connectives for classical dialectical logic, classical tensor product $\otimes$ and classical tensor sum ∇, are definable in terms of the Heyting tensor product $\circ$ and tensor negation $\neg$. For any two $\circ$-composable terms $z \xrightarrow{s} y$ and $y \xrightarrow{r} x$ the tensor product term $z \xrightarrow{s \otimes r} x$ and the tensor sum term $z \xrightarrow{s \nabla r} x$ are $\neg\neg$-closed terms define by $s \otimes r \stackrel{\text{df}}{=} \neg\neg(s \circ r)$ and $s \nabla r \stackrel{\text{df}}{=} \neg(\neg r \otimes \otimes \neg s) = \neg(\neg r \circ \neg s)$. For all terms we immediately have the DeMorgans laws $\neg(s \nabla r) = \neg r \otimes \neg s$ and $\neg(s \oplus r) = \neg s \Delta \neg r$, for $\mathcal{Z}(\mathbf{H})$-open terms we have the DeMorgans inequalities $\neg(s \otimes r) \preceq \neg r \nabla \neg s$ and $\neg(s \Delta r) \preceq \neg s \oplus \neg r$, and for $\neg\neg$-closed terms we have the DeMorgans laws $\neg(s \otimes r) = \neg r \nabla \neg s$ and $\neg(s \Delta r) = \neg s \oplus \neg r$.

A Heyting term is *polar* when it is $\neg\neg$-closed and $\mathcal{Z}(\mathbf{H})$-open; that is, when the term is in $\neg\neg\mathcal{Z}(\mathbf{H})$. The *pole* of any Heyting term is the double negation of its $\mathcal{Z}(\mathbf{H})$-interior (if it exists). The lax functoriality of double negation $\neg\neg$ implies that the classical tensor product is associative $t \otimes (s \otimes \otimes r) = (t \otimes s) \otimes r$ on polar terms. Also, types are identities $y \otimes r = r = r \otimes x$ on polar terms. The *Boolean pole* of $\mathcal{Z}(\mathbf{H})$, denoted by $\mathcal{Z}(\mathbf{H})^{\oplus}_{\otimes}$, is the join bisemilattice $\mathcal{Z}(\mathbf{H})^{\oplus}_{\otimes} = \langle\langle \neg\neg\mathcal{Z}(\mathbf{H}), \preceq, \otimes, \mathrm{Id}\rangle, \oplus, 0\rangle$ consisting of all types and all polar terms (join bisemilattice since finite homset joins exist, but not necessarily finite homset meets), with the classical tensor product and Boolean sum. $\mathcal{Z}(\mathbf{H})^{\oplus}_{\otimes}$ is a lax (Heyting) subcategory of $\mathcal{Z}(\mathbf{H})$. Dually, a Heyting term is *antipolar* when it is $\neg\neg$-closed and $\mathcal{Z}(\mathbf{H})$-closed; that is, when it is the tensor negation of a polar term. The image $\neg\mathcal{Z}(\mathbf{H})$ of tensor negation on the pole is the collection of all antipolar terms. The tensor DeMorgans laws (and the associativity of the tensor product $\otimes$) imply that the classical tensor sum ∇ is associative $t \nabla (s \nabla r) = (t \nabla s) \nabla r$ on antipolar terms. Also, types are identities $y \nabla r = r = r \nabla x$ on antipolar terms. The *Boolean antipole* of $\mathcal{Z}(\mathbf{H})$, denoted by $\mathcal{Z}(\mathbf{H})^{\Delta}_{\nabla}$, is the meet bisemilattice $\mathcal{Z}(\mathbf{H})^{\Delta}_{\nabla} = \langle\langle \neg\mathcal{Z}(\mathbf{H}), \preceq, \nabla, \mathrm{Id}\rangle, \Delta, 1\rangle$ consisting of all types and all antipolar terms, and the classical tensor sum and Boolean product. Moreover, tensor negation is a 2-involution, a morphism of join bisemilattices $\mathcal{Z}(\mathbf{H})^{\oplus}_{\otimes} \xrightarrow{\neg} \mathcal{Z}(\mathbf{H})^{\Delta\,\mathrm{coop}}_{\nabla}$ and a morphism of meet bisemi-

lattices $\mathcal{Z}(\mathbf{H})^{\oplus^{\mathrm{coop}}}_{\otimes} \rightleftarrows \mathcal{Z}(\mathbf{H})^{\triangle}_{\nabla}$: $\neg$ is self-inverse $\neg\neg r = r$, $\neg x = x$, $\neg$ switches source and target $\neg(y \xrightarrow{r} x) = x \xrightarrow{\neg r} y$, and $\neg$ is (contravariant) monotonic on homsets $r \preceq_{y,x} s$ implies $\neg s \preceq_{x,y} \neg r$. This complex, consisting of a join and meet bisemilattice and the negation involution between them, is called the *Boolean* of $\mathcal{Z}(\mathbf{H})$ or the *Boolean center* of $\mathbf{H}$, and is denoted by $\mathcal{B}(\mathcal{Z}(\mathbf{H}))$.

The special property $s \perp_{\otimes} r$ iff $s \preceq \neg r$ called the *orthogonality-entailment axiom*, which relates term-orthogonality with term-order, holds for all polar terms. Equivalently, the special property $s \perp^{\mathrm{co}}_{\nabla} r$ iff $\neg s \preceq r$, which relates term-coorthogonality with term-order, holds for all antipolar terms. The Boolean center $\mathcal{B}(\mathcal{Z}(\mathbf{H}))$ is quasisymmetric: the Boolean pole $\mathcal{Z}(\mathbf{H})^{\oplus}_{\otimes}$ is a quasisymmetric category since a Heyting term $y \xrightarrow{r} x$ is $\circ$-quasisymmetric iff it is $\otimes$-quasisymmetric, and the Boolean antipole $\mathcal{Z}(\mathbf{H})^{\triangle}_{\nabla}$ is a coquasisymmetric category since a Heyting term $y \xrightarrow{r} x$ being $\circ$-coquasisymmetric implies that it is ∇-coquasisymmetric. For any pair of terms in either the pole or the antipole of the Boolean center, the Heyting tensor product and the classical tensor connectives are arranged as $s \circ r \preceq s \otimes r \preceq s \nabla r$. When $\mathbf{H}$ is quasisymmetric the Boolean center $\mathcal{B}(\mathbf{H})$ consists of all $\neg\neg$-closed terms.

A *polarized bisemilattice* $\mathbf{P}$ consists of two bisemilattices, a join bisemilattice $\mathbf{P}^{\oplus}_{\otimes} = \langle\langle \mathbf{P}^{\oplus}_{\otimes}, \preceq_{\otimes}, \otimes, \mathrm{Id}\rangle, \oplus, 0\rangle$ and a meet bisemilattice $\mathbf{P}^{\triangle}_{\nabla} = \langle\langle \mathbf{P}^{\triangle}_{\nabla}, \preceq_{\nabla}, \nabla, \mathrm{Id}\rangle, \Delta, 1\rangle$, called the *pole* and *antipole* of $\mathbf{P}$ respectively, and two morphisms of bisemilattices, a morphism of join bisemilattices $\mathbf{P}^{\oplus}_{\otimes} \rightarrow \mathbf{P}^{\triangle^{\mathrm{coop}}}_{\nabla}$ and a morphism of meet bisemilattices $\mathbf{P}^{\oplus^{\mathrm{coop}}}_{\otimes} \leftarrow \mathbf{P}^{\triangle}_{\nabla}$ which are inverse $\neg \cdot \cdot \neg^{\mathrm{coop}} = \mathrm{Id}$ to each other. Just as for Heyting categories, objects and arrows in either the pole $\mathbf{P}^{\oplus}_{\otimes}$ or the antipole $\mathbf{P}^{\triangle}_{\nabla}$ are called *types* and *terms*, respectively. The Boolean center $\mathcal{B}(\mathcal{Z}(\mathbf{H}))$ of any Heyting category $\mathbf{H}$ is a polarized bisemilattice. Morphisms of polarized bisemilattices can be defined in either a polar or an antipolar sense. A *morphism of polarized bisemilattices* $\mathbf{P} \xrightarrow{H} \mathbf{Q}$ consists of a morphism of join bisemilattices $\mathbf{P}^{\oplus}_{\otimes} \xrightarrow{H^{\oplus}_{\otimes}} Q^{\oplus}_{\otimes}$ called the *pole* of H, and a morphism of meet bisemilattices $\mathbf{P}^{\triangle}_{\nabla} \xrightarrow{H^{\triangle}_{\nabla}} Q^{\triangle}_{\nabla}$ called the *antipole* of H, which are interdefinable with $H^{\triangle}_{\nabla} \stackrel{\mathrm{df}}{=} \neg_P \cdot (H^{\oplus}_{\otimes})^{\mathrm{coop}} \cdot (\neg_Q)^{\mathrm{coop}}$ and $H^{\oplus}_{\otimes} \stackrel{\mathrm{df}}{=} \neg_P \cdot \cdot (H^{\triangle}_{\nabla})^{\mathrm{coop}} \cdot (\neg_Q)^{\mathrm{coop}}$.

Boolean categories. Ignoring idempotency and commutativity, a Boolean algebra $B = \langle B, \leqq, \wedge, \vee, 1, 0, \neg\rangle$ can be viewed as two monoidal semilattices, a monoidal join semilattice $B^{\vee}_{\wedge} = \langle\langle B, \leqq, \wedge, 1\rangle, \vee, 0\rangle$ and a monoidal meet semilattice $B^{\wedge}_{\vee} = \langle\langle B, \leqq, \vee, 0\rangle, \wedge, 1\rangle$ on an underlying poset $\langle B, \leqq\rangle$ with negation $\neg$ being an internal involution: a monoidal join semilattice morphism $B^{\vee}_{\wedge} \rightarrow B^{\wedge^{\mathrm{coop}}}_{\vee}$, $b \leqq b'$ implies $\neg b' \leqq \neg b$, $\neg(c \wedge b) = (\neg c) \vee (\neg b)$, $\neg 1 = 0$, $\neg(b \vee \vee b') = (\neg b) \wedge (\neg b')$ and $\neg 0 = 1$, and a monoidal meet semilattice morphism $B^{\vee^{\mathrm{coop}}}_{\wedge} \leftarrow B^{\wedge}_{\vee}$, which is self-inverse $\neg(\neg b) = b$ or $\neg \cdot \neg^{\mathrm{coop}} = \mathrm{Id}$. More general-

ly, a *Boolean category* $\mathbf{B}$ is a polarized bisemilattice for which the term-sets, type-sets and homset-order of the pole and the antipole coincide $\mathrm{Ar}(\mathbf{B}) = \mathrm{Ar}(\mathbf{B}^{\oplus}_{\otimes}) = \mathrm{Ar}(\mathbf{B}^{\Delta}_{\nabla})$, $\mathrm{Obj}(\mathbf{B}) = \mathrm{Obj}(\mathbf{B}^{\oplus}_{\otimes}) = \mathrm{Obj}(\mathbf{B}^{\Delta}_{\nabla})$ and $\preceq_{\otimes} = \preceq_{\nabla} = \preceq$ (and are not just isomorphic as in polarized bisemilattices, where the term-sets and type-sets are not identical, but only in bijective correspondence via negation), and which satisfies the *orthogonality-entailment axiom*

$$s \perp_{\otimes} r \quad \text{iff} \quad s \preceq \neg r$$

for all opposed terms $y \xrightarrow{r} x$ versus $y \xleftarrow{s} x$, which relates term-orthogonality with term-order (because of the precise duality expressed through tensor negation, $s \perp_{\otimes} r$ iff $\neg r \perp^{\mathrm{co}}_{\nabla} \neg s$, polar orthogonality can be expressed as, and is equivalent to, antipolar coorthogonality).

In more detail, a Boolean category $\mathbf{B}$ consists of a set of types (objects) $\mathrm{Type}(\mathbf{B})$, a set of terms (arrows) $\mathrm{Term}(\mathbf{B})$ ordered type-wise by a partial order $\preceq$ which has homset lattice join $\oplus$ and homset lattice meet Δ and two category compositions $\otimes$ and ∇, where the pole $\mathbf{B}^{\oplus}_{\otimes} = \langle\langle \mathbf{B}, \preceq, \otimes, \mathrm{Id}\rangle, \oplus, 0\rangle$ and the antipole $\mathbf{B}^{\Delta}_{\nabla} = \langle\langle \mathbf{B}, \preceq, \nabla, \mathrm{Id}\rangle, \Delta, 1\rangle$ are join and meet bisemilattices, respectively, with an internal 2-involution $\mathbf{B}^{\oplus}_{\otimes} \rightleftarrows \mathbf{B}^{\Delta\,\mathrm{coop}}_{\nabla}$. A Boolean category is finitely distributive in two senses: from the left $s \otimes (\oplus_i r_i) = \oplus_i (s \otimes r_i)$ in $\mathbf{B}^{\oplus}_{\otimes}$ and $s \nabla (\Delta_i r_i) = \Delta_i (s \nabla r_i)$ in $\mathbf{B}^{\Delta}_{\nabla}$, and also from the right in both poles. The tensor negation is (1) a doubly-contravariant (everything "flips") morphism of join bisemilattices $\mathbf{B}^{\oplus}_{\otimes} \rightleftarrows \mathbf{B}^{\Delta\,\mathrm{coop}}_{\nabla}$ identity on types, $\neg(y \xrightarrow{r} x) = x \xrightarrow{\neg r} y$, $\neg(s \oplus r) = (\neg r) \nabla (\neg s)$, $\neg x = x$, $r \preceq_{y,x} r'$ implies $\neg r' \preceq_{x,y} \neg r$ and $\neg(r \oplus r') = (\neg r) \Delta (\neg r')$; (2) a doubly-contravariant morphism of meet bisemilattices $\mathbf{B}^{\oplus\,\mathrm{coop}}_{\otimes} \rightleftarrows \mathbf{B}^{\Delta}_{\nabla}$ in the reverse direction and opposite sense, $\neg(s \nabla r) = (\neg r) \otimes (\neg s)$ and $\neg(r \Delta r') = (\neg r) \oplus (\neg r')$; (3) which is self-inverse $\neg(\neg r) = r$. In a Boolean category orthogonality preserves composition, in the sense that: $q \perp s$ and $p \perp r$ implies $(p \otimes q) \perp (s \otimes r)$. Also, a Boolean category satisfies the product-sum comparison (or "mix") axiom: $s \otimes r \preceq_{z,x} s \nabla r$ for all terms $z \xrightarrow{s} y$ and $y \xrightarrow{r} x$. A one object Boolean category is called a *Boolean monoid*. The homsets $\mathbf{B}[x, x]$ are Boolean monoids for each type x. A Boolean category is *complete* when the poles are both complete Heyting categories; that is, the homsets are complete lattices, tensor product is completely distributive (continuous) w.r.t. Boolean sum, and tensor sum is completely distributive (continuous) w.r.t. Boolean product. Morphisms of Boolean categories are just morphisms of polarized bisemilattices.

A term $y \xrightarrow{r} x$ in a Boolean category is *invertible* when its tensor negation is a categorical inverse: $\neg r \otimes r = x$ and $r \otimes \neg r = y$. So invertible terms are the same as B-isomorphisms. For isomorphisms the direct and inverse image operators are isomorphisms of Boolean monoids. Clearly, all identities are isomorphisms. Isomorphisms are closed under tensor product, tensor sum and tensor negation. In fact, the tensor sum collapses to the tensor

product $s \nabla r = s \otimes r$ for composable isomorphisms. When all terms in a Boolean category are isomorphisms, the Boolean category is known as a *lattice-ordered groupoid.* In general, the collection of all isomorphisms in a Boolean category **B** is a Boolean subcategory of **B** which is a lattice-ordered groupoid. A summary of the appropriate semantic domains for various logics is given in Figure 1.

	Intuitionistic	**Classical**
Standard logic	Heyting algebras (in particular, subset algebras)	Boolean algebras
Linear logic	commutative Heyting monoids (in particular, "phase spaces")	commutative Boolean monoids
Dialectical logic (this paper)	Heyting categories (in particular, subset categories)	(quasisymmetric) Boolean categories
Dialectical logic (extended version)	Heyting categories with type sums (in particular, distributor categories)	(quasisymmetric) Boolean categories with type sums

Figure 1. Semantic domains for various logics

Recall that a term $y \xrightarrow{r} x$ is $\mathbf{B}^{\oplus}_{\otimes}$-quasisymmetric when $p \otimes r \preceq x$ iff $r \otimes p \preceq y$, and is $\mathbf{B}^{\triangle}_{\nabla}$-coquasisymmetric when $p \nabla r \succeq x$ iff $r \nabla p \succeq y$. So r is $\mathbf{B}^{\oplus}_{\otimes}$-quasisymmetric iff $\neg r$ is $\mathbf{B}^{\triangle}_{\nabla}$-coquasisymmetric. This means that the tensor negation 2-involution restricts and corestricts precisely to the center of $\mathbf{B}^{\oplus}_{\otimes}$ and the cocenter of $\mathbf{B}^{\triangle}_{\nabla}$: $\mathcal{Z}(\mathbf{B}^{\oplus}_{\otimes}) \xrightarrow{\neg} \mathcal{Z}(\mathbf{B}^{\triangle}_{\nabla})^{\mathrm{coop}}$. Call this the *center* of **B**, and denote it by $\mathcal{Z}(\mathbf{B})$. A Boolean category **B** is *quasisymmetric* when $\mathcal{Z}(\mathbf{B}) = \mathbf{B}$. Quasisymmetric Boolean categories (and the Boolean center of their associated closed subset categories) are fundamental semantic structures for complete classical dialectical logic.

Let $y \xrightarrow{r} x$ be any fixed $\mathbf{B}^{\otimes}_{\oplus}$-term. For any $\mathbf{B}^{\nabla}_{\Delta}$-term $y \xrightarrow{t} z$ with source type in common with r, define the *left tensor implication* $\mathbf{B}^{\nabla}_{\Delta}$-term $x \xrightarrow{r \backslash t} z$ by $r \backslash t \stackrel{\mathrm{df}}{=} \neg r \nabla t$. Similarly, for any $\mathbf{B}^{\nabla}_{\Delta}$-term $z \xrightarrow{s} x$ with target type in common with r define the *right tensor implication* $\mathbf{B}^{\nabla}_{\Delta}$-term $z \xrightarrow{s / r} y$ by $s / r \stackrel{\mathrm{df}}{=} s \nabla \neg r$. The dialectical axioms $t \otimes r \preceq_{z,x} s$ iff $t \preceq_{z,y} s / r$ and $r \otimes s \preceq_{y,z} t$ iff $s \preceq_{x,z} r \backslash t$ hold on quasisymmetric terms. Adjoining these implication operators to the center pole $\mathcal{Z}(\mathbf{B}^{\otimes}_{\oplus})$ makes this into a quasisymmetric Heyting category $\mathcal{H}(\mathcal{Z}(\mathbf{B}))$ called the *Heyting center* of **B**, whose tensor negation is the same as in **B**. So all terms in $\mathcal{H}(\mathcal{Z}(\mathbf{B}))$ are double negation closed.

THEOREM 1 (Center Reflection). *If* **H** *is a quasisymmetric Heyting category, then the Boolean center* $\mathcal{B}(\mathbf{H})$ *is a quasisymmetric Boolean category.*

Any quasisymmetric Boolean category **B** *is a quasisymmetric Heyting category* $\mathcal{H}(\mathbf{B})$. *For any quasisymmetric Boolean category* **B**, *the Boolean center of* **B** *as a Heyting category is just* **B** *itself* $\mathcal{B}(\mathcal{H}(\mathbf{B})) = \mathbf{B}$. *For any quasisymmetric Heyting category* **H**, *the Boolean center as a Heyting category, is just the center pole* $\mathcal{H}(\mathcal{B}(\mathbf{H})) = \mathbf{H}^{\oplus}_{\otimes}$, *the lax subHeyting category of* **H** *consisting of double negation closed terms.*

3. Classical axiomatics

We follow both the semantics of dialectical processes and the axiomatics given by Girard for linear logic. However, when linear logic deviates from dialectical process semantics, we follow the latter. A hallmark of both dialectical and linear logic is the fact that the standard connectives and truth-values split into tensors and Booleans, as in Table 1.

Standard logic	Dialectical logic	Uses
$\wedge$ Boolean product	$\otimes_{z,y,x}$ tensor (horizontal) product	direct flow
	$\Delta_{y,x}$ Boolean (vertical) product	parallelism & inverse flow
$\top$ true	$\langle m, x\rangle$ monoids (comonoids)	tensor validity
	$1_{y,x}$ top process	Boolean validity
$\vee$ Boolean sum	$\nabla_{z,y,x}$ tensor (horizontal) sum	inverse flow
	$\oplus_{y,x}$ Boolean (vertical) sum	parallelism & direct flow
$\perp$ false	$\langle m, x\rangle$ monoids (comonoids)	orthogonality
	$0_{y,x}$ bottom process	disjointness

Table 1. Splitting of connectives and truth values

Language. There is a collection of *type symbols* $x, y, z, \cdots$, and a collection of *atoms* or *atomic term symbols* $a, b, c, \cdots$. Each atom a is a term formula, and has a unique source type y and a unique target type x, denoted by $y \xrightarrow{a} x$. Each atom $y \xrightarrow{a} x$ has a *dual* or *complement* $x \xrightarrow{\dot{a}} y$. Atoms and their duals are called *literals*. So type symbols are the nodes of a graph **Lang**, and literals (and other composite term formulas) form the edges. For each pair of types y and x, there are two distinguished term symbols $y \xrightarrow{0} x$ and $y \xrightarrow{1} x$. Each type x is represented as a term formula $x \xrightarrow{x} x$, which is a self-loop at node x in the graph **Lang**. Composite term formulas are built up recursively from literals by horizontally applying the tensor operation symbols $\otimes$ and ∇, and vertically applying the Boolean operation symbols $\oplus$ and Δ, in an obvious type-consistent fashion. Term formulas are also called terms. This will be legitimized below when it is shown that the (equivalence classes of) term formulas form a Boolean category. Following Girard's approach, there is an

external involution $\mathbf{Lang} \xrightarrow{\neg} \mathbf{Lang}^{\mathrm{op}}$ called *tensor negation*, which is defined recursively on terms as follows: **base** $\neg a \stackrel{\mathrm{df}}{=} \dot{a}$ and $\neg(\dot{a}) \stackrel{\mathrm{df}}{=} a$; **recursion** $\neg x \stackrel{\mathrm{df}}{=} x$, $\neg(\beta \otimes \alpha) \stackrel{\mathrm{df}}{=} (\neg\alpha)\nabla(\neg\beta)$ and $\neg(\beta \nabla \alpha) \stackrel{\mathrm{df}}{=} (\neg\alpha) \otimes (\neg\beta)$, $\neg(\alpha \oplus \alpha') \stackrel{\mathrm{df}}{=} (\neg\alpha)\Delta(\neg\alpha')$ and $\neg(\alpha \Delta \alpha') \stackrel{\mathrm{df}}{=} (\neg\alpha) \oplus (\neg\alpha')$, and $\neg(y \xrightarrow{0} x) \stackrel{\mathrm{df}}{=} x \xrightarrow{1} y$ and $\neg(y \xrightarrow{1} x) \stackrel{\mathrm{df}}{=} x \xrightarrow{0} y$.

FACT 1. $\neg(\neg\alpha) = \alpha$ *for every term* α.

In addition to the previous symbols which specify types and terms, there are two special symbols $\vdash$ and $\perp$ which specify the binary relation of *entailment* between parallel terms and the binary relation of *orthogonality* between opposed terms, respectively. The entailment and orthogonality relations on terms give two equivalent ways in which to specify dialectical logic.

Inference rules. The formal semantics of classical dialectical logic will be defined via axioms and inference rules. The novelty of this approach lies in the use of orthogonality assertions, rather than just term entailment assertions alone. An orthogonality assertion is a statement of the form $\beta \perp \alpha$ for two opposed terms $y \xrightarrow{\alpha} x$ versus $y \xleftarrow{\beta} x$, and when $\beta \perp \alpha$ holds, we say that α is *orthogonal* to β. An orthogonality assertion is interpreted as the orthogonality of the terms specified by the opposed term formulas. The orthogonality relation $\perp$ has a negation-dual relation $\perp^{\mathrm{co}}$, called *coorthogonality*, and defined by $\beta \perp^{\mathrm{co}} \alpha$ when $\neg\alpha \perp \neg\beta$. An entailment assertion is a statement of the form $\alpha \vdash \beta$ for two parallel terms $y \xrightarrow{\alpha,\beta} x$, and when $\alpha \vdash \beta$ holds, we say that α *entails* β. The entailment relation $\vdash$ has an obvious dual relation $\vdash^{\mathrm{op}}$ defined by $\beta \vdash^{\mathrm{op}} \alpha$ when $\alpha \vdash \beta$; so that, $\vdash^{\mathrm{op}} = \dashv$. We use the equivalence notation $\alpha \vdash\!\dashv \beta$ when both $\alpha \vdash \beta$ and $\beta \vdash \alpha$ hold, and we say that α is *entailment equivalent* to β. When "α entails identity", that is when $\alpha \vdash x$ holds, we say that the term α itself is *provable*. So an endoterm $x \xrightarrow{\alpha} x$ is provable iff $\alpha \in {\downarrow}(x)$ the principal ideal of the identity term.

We give two versions of inference rules for the term calculus: an *entailment version* which is closely related to the semantics of dialectical logic, and an *orthogonality version* which extends Girard's version [Girard] of the linear logic. In each version we group the rules according to their semantics: the vertical aspect in Table 2 and the horizontal aspect in Table 3. The homset-order axioms in the two versions are immediately equivalent; in fact, the logical axioms are equivalent to reflexivity of entailment, the cut rule is equivalent to transitivity of entailment, and symmetry is equivalent to contravariance of tensor negation. So entailment is a homset preorder on terms, and **Lang** is a preordered graph. Similarly, the tensor axioms, the $\otimes\nabla$-rule and monotonicity of tensor product $\otimes$, are equivalent. By applying tensor negation, the monotonicity of tensor product $\otimes$ and the monotonicity of tensor sum ∇ are equivalent facts. The cut rule implies that orthogonality is monotonic: if $\beta \perp \alpha$ and $\alpha' \vdash \alpha$ then $\beta \perp \alpha'$. The Boolean rules assert that $\oplus$ is a least upper bound and that Δ is a greatest lower bound in the entailment order. The zero rule provides the axiomatics for both bottom 0 and top 1.

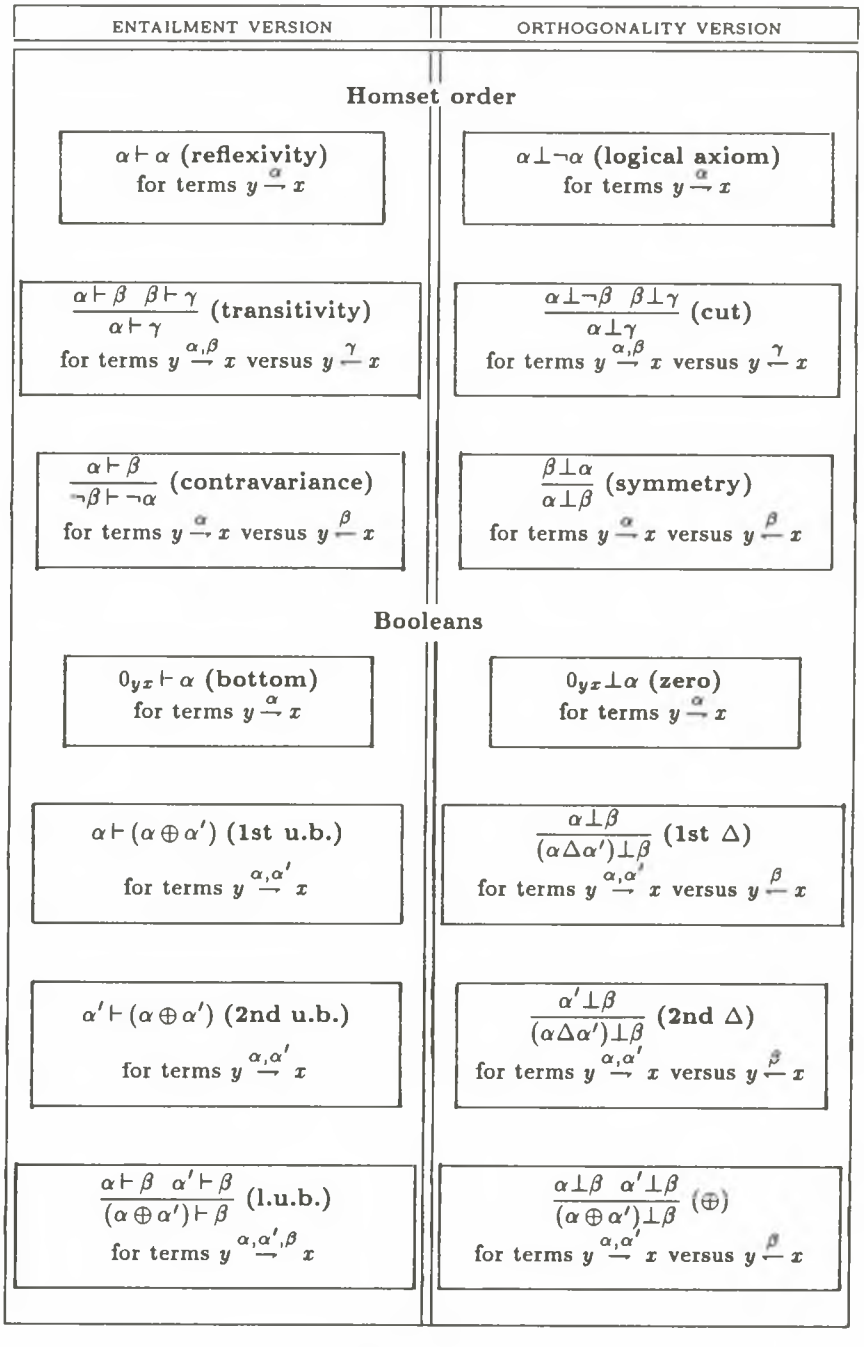


Table 2. Vertical aspect of term rules

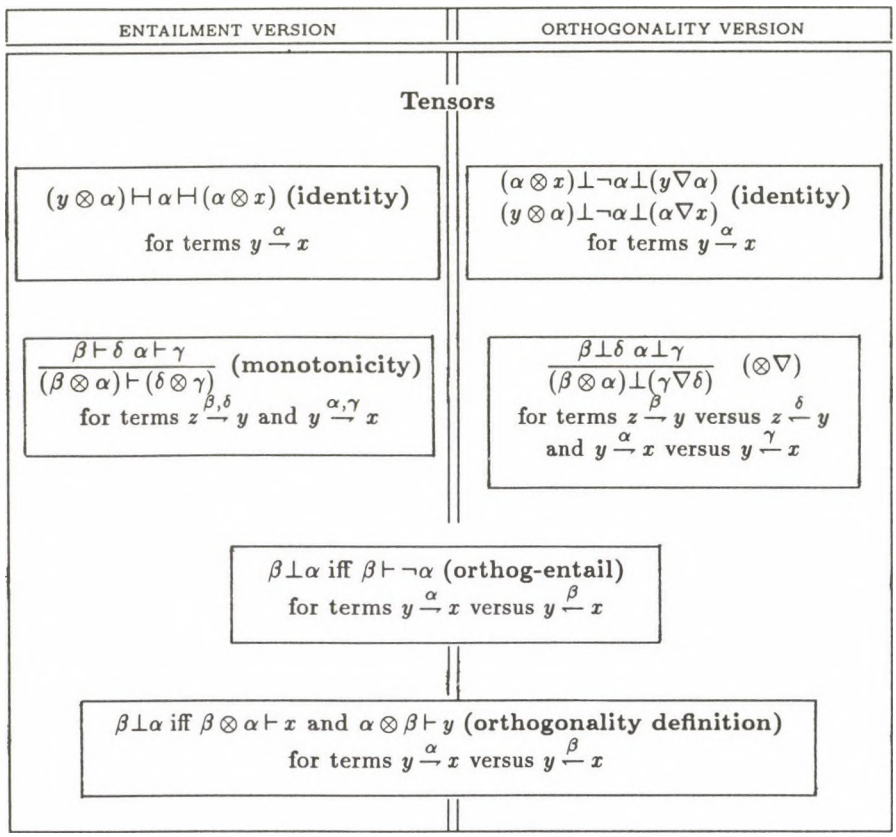


Table 3. Horizontal aspect of term rules

Thus, the (internal) vertical aspect of term formulas has the structure of a lattice; with the (external) tensor negation, ignoring types, it has the structure of a Boolean algebra. The entailment axioms, minus contravariance, are essentially the axioms for a join bisemilattice. The vertical aspect of the basic calculus corresponds to standard (propositional) logic. The horizontal aspect of the basic calculus, minus the orthogonality definition axiom, is a dialectical logic analog or typed version of the "multiplicative fragment" adjoined by linear logic. The definition of orthogonality, which axiomatizes "Boolean orthogonality" or the definition of orthogonality in Boolean categories, separates dialectical logic from typed linear logic. We want to show that the horizontal aspect of term formulas has categorical structure for both tensor product and tensor sum. We can do this quite simply by extending entailment to sequences of term formulas.

Sequents. A *sequent* α is a path of term formulas (**Lang**-edges) $y \xrightarrow{\alpha} x =$

$= y \xrightarrow{\alpha_n} x_{n-1} \rightharpoonup \cdots \rightharpoonup x_1 \xrightarrow{\alpha_1} x$. Such a path is a typed version of a sequence of term formulas. The concatenation of two sequents $z \xrightarrow{\beta} y$ and $y \xrightarrow{\alpha} x$ is denoted by $z \xrightarrow{\beta \circ \alpha} x$. The empty sequent at type symbol x is denoted by $x \xrightarrow{\varepsilon_x} x$. So sequents are arrows in a free (path) category **Lang*** having concatenation $\circ$ as composition and empty paths ε_x as identities. The category of sequents **Lang*** inherits from the graph of terms **Lang** a weak *vector entailment* homset order $\vec{\vdash}$, defined by $\alpha \vec{\vdash} \beta$ when $|\alpha| = |\beta|$ and $\alpha_i \vdash \beta_i$ for all $1 \leqq i \leqq n$, where $\alpha = \alpha_n \circ \cdots \circ \alpha_1$. Clearly, sequent concatenation is monotonic w.r.t. vector entailment: if $\beta \vec{\vdash} \delta$ and $\alpha \vec{\vdash} \gamma$ then $(\beta \circ \alpha) \vec{\vdash} (\delta \circ \gamma)$ for any two composable parallel pairs of sequents $z \xrightarrow{\beta,\delta} y$ and $y \xrightarrow{\alpha,\gamma} x$. So $\vec{\textbf{Lang}}^* \stackrel{\text{df}}{=} (\textbf{Lang}^*, \vec{\vdash})$ is a bipreorder (preordered category). Extend tensor negation to sequents by defining the sequent "vector" *tensor negation* $\vec{\neg}\alpha \stackrel{\text{df}}{=} \neg\alpha_1 \circ \cdots \circ \neg\alpha_n$ for any sequent $y \xrightarrow{\alpha} x$ which is the path of terms $\alpha = \alpha_n \circ \cdots \circ \alpha_1$; in particular, $\vec{\neg}\varepsilon_x \stackrel{\text{df}}{=} \varepsilon_x$. Vector tensor negation is contravariant: if $\alpha \vec{\vdash} \beta$ then $\vec{\neg}\beta \vec{\vdash} \vec{\neg}\alpha$. So vector tensor negation is a categorical involution $\vec{\neg}\vec{\neg}\alpha = \alpha$; that is, a contravariant functor $\vec{\textbf{Lang}}^* \xrightarrow{\vec{\neg}} (\vec{\textbf{Lang}}^*)^{\text{coop}}$, which is self-inverse $\vec{\neg} \cdot (\vec{\neg})^{\text{coop}} = \text{Id}$. The category of sequents, vector entailment, and vector tensor negation form a polarized bipreorder $\vec{\textbf{Lang}}^*$.

Sequents will be interpreted in Boolean categories. A sequent can be interpreted in a Boolean category in either a polar sense (using $\otimes$) or an antipolar sense (using ∇). The two senses are inter-translatable via tensor negation. In Girard's version of linear logic, sequents are interpreted in the antipolar sense. The interpretation of a sequent $y \xrightarrow{\alpha} x$ in the polar sense is done via the *tensor product term* $y \xrightarrow{\otimes(\alpha)} x$, a sequent of length one, which is defined by $\otimes(\alpha) \stackrel{\text{df}}{=} \alpha_n \otimes \cdots \otimes \alpha_1$. More precisely, **base** $\otimes(\varepsilon_x) \stackrel{\text{df}}{=} x$ for any type x, and **induction** $\otimes(\beta \circ \alpha) \stackrel{\text{df}}{=} \beta \otimes \otimes(\alpha)$ for any term $z \xrightarrow{\beta} y$ and any sequent $y \xrightarrow{\alpha} x$. In particular, $\otimes(\alpha) = \alpha \otimes x$ for any term $y \xrightarrow{\alpha} x$. So the tensor product operator is a type-preserving graph morphism $\textbf{Lang}^* \xrightarrow{\otimes} \textbf{Lang}$ from the category of sequents **Lang*** to the graph of terms **Lang**. Dually, the interpretation of a sequent $y \xrightarrow{\alpha} x$ in the antipolar sense is done via the *tensor sum term* $y \xrightarrow{\nabla(\alpha)} x$, a sequent of length one, which is defined by $\nabla(\alpha) \stackrel{\text{df}}{=} \alpha_n \nabla \cdots \nabla \alpha_1$. More precisely, **base** $\nabla(\varepsilon_x) \stackrel{\text{df}}{=} x$ for any type x, and **induction** $\nabla(\beta \circ \alpha) \stackrel{\text{df}}{=} \nabla(\beta)\nabla\alpha$ for any sequent $z \xrightarrow{\beta} y$ and any term $y \xrightarrow{\alpha} x$. In particular, $\nabla(\alpha) = x \nabla \alpha$ for any term α. So the tensor sum operator is also a type-preserving graph morphism $\textbf{Lang}^* \xrightarrow{\nabla} \textbf{Lang}$. By induction we can show that the tensor product and tensor sum operations are related by

the DeMorgan's laws $\neg(\otimes\alpha) = \nabla(\vec{\neg}\alpha)$ and $\neg(\nabla\alpha) = \otimes(\vec{\neg}\alpha)$.

In the polar sense of interpretation, we require that each sequent α be logically equivalent to its tensor product term $\otimes(\alpha)$. So define a *polar entailment* homset order $\vdash_\otimes$ by $\alpha \vdash_\otimes \beta$ when $\otimes(\alpha) \vdash \otimes(\beta)$. Polar entailment partially orders **Lang***-homsets, if we quotient out by logical equivalence $\dashv\vdash_\otimes$ defined by: $\alpha \dashv\vdash_\otimes \beta$ when both $\alpha \vdash_\otimes \beta$ and $\beta \vdash_\otimes \alpha$ hold. Then any sequent $y \xrightarrow{\alpha} x$ is entailment equivalent to its associated tensor product term $\alpha \dashv\vdash_\otimes \otimes(\alpha)$, as is required by the polar interpretation, since $\otimes(\otimes(\alpha)) = \otimes(\alpha) \otimes x \dashv\vdash \otimes(\alpha)$. The tensor product of terms is associative, up to polar entailment equivalence (for sequents), since $\gamma \otimes (\beta \otimes \alpha) \dashv\vdash_\otimes \gamma \circ (\beta \circ \alpha) = (\gamma \circ \beta) \circ \alpha \dashv\vdash_\otimes (\gamma \otimes \beta) \otimes \alpha$. Polar entailment equivalence $\dashv\vdash_\otimes$ extends term entailment equivalence $\dashv\vdash$; that is, polar entailment equivalence coincides with entailment equivalence on terms, $\beta \dashv\vdash_\otimes \alpha$ iff $\beta \dashv\vdash \alpha$ for all terms $y \xrightarrow{\alpha,\beta} x$. So, the tensor product of terms is associative, up to term entailment equivalence: $\gamma \otimes (\beta \otimes \alpha) \dashv\vdash (\gamma \otimes \beta) \otimes \alpha$. By induction tensor product preserves composition, up to term equivalence $\otimes(\beta \circ \alpha) \dashv\vdash \otimes(\beta) \otimes \otimes(\alpha)$. Sequent concatenation is monotonic w.r.t. polar entailment: if $\beta \vdash_\otimes \delta$ and $\alpha \vdash_\otimes \gamma$ then $(\beta \circ \alpha) \vdash_\otimes (\delta \circ \gamma)$ for any two composable parallel pairs of sequents $z \xrightarrow{\beta,\delta} y$ and $y \xrightarrow{\alpha,\gamma} x$, since tensor product is monotonic. So, the category of sequents **Lang*** forms a bipreorder $\mathbf{Lang}^*_\otimes \stackrel{\text{df}}{=} \langle \mathbf{Lang}^*, \vdash_\otimes \rangle$ with polar entailment $\vdash_\otimes$. By induction using the monotonicity rule, the tensor product operator is monotonic w.r.t. vector entailment: if $\alpha \vec{\vdash} \beta$ then $\otimes(\alpha) \vdash \otimes(\beta)$. So vector entailment is weaker than polar entailment: if $\alpha \vec{\vdash} \beta$ then $\alpha \vdash_\otimes \beta$.

Dually, in the antipolar sense of interpretation, we require that each sequent α be logically equivalent to its tensor sum term $\nabla(\alpha)$. So define an *antipolar entailment* homset order $\vdash_\nabla$ by $\alpha \vdash_\nabla \beta$ when $\nabla(\alpha) \vdash \nabla(\beta)$. The category of sequents **Lang*** forms a bipreorder $\mathbf{Lang}^*_\nabla \stackrel{\text{df}}{=} \langle \mathbf{Lang}^*, \vdash_\nabla \rangle$ with antipolar entailment $\vdash_\nabla$. Again, vector entailment is weaker than antipolar entailment: if $\alpha \vec{\vdash} \beta$ then $\alpha \vdash_\nabla \beta$. The polar and antipolar orders are two alternate interpretations for the entailment relation $\vdash$ on sequents. They are polar duals, and are interdefinable via the equivalence: $\alpha \vdash_\otimes \beta$ iff $\vec{\neg}\beta \vdash_\nabla \vec{\neg}\alpha$. More concisely, vector tensor negation is an involution $\mathbf{Lang}^*_\otimes \xrightarrow{\vec{\neg}} (\mathbf{Lang}^*_\nabla)^{\text{coop}}$. So the category of sequents, the two polarities of entailment, and vector tensor negation form a polarized bipreorder **Lang***.

The term category. Entailment partially orders **Lang**-homsets, if we quotient out by logical equivalence $\dashv\vdash$. Entailment equivalence quotienting is done automatically when we use the closed subset construction. For any term $y \xrightarrow{\alpha} x$, let $[y] \xrightarrow{[\alpha]} [x]$ denote the quotient term (entailment equivalence class) of α. Let **Term** denote the quotient graph of **Lang**; that is, **Term** is the graph of types and quotient terms. Define the Boolean and tensor operations

on quotient terms via representatives. For example, define the tensor product and tensor sum of quotient terms by $[\beta] \otimes [\alpha] \stackrel{\text{df}}{=} [\beta \otimes \alpha]$ and $[\beta] \nabla [\alpha] \stackrel{\text{df}}{=} [\beta \nabla \alpha]$. Define the quotient entailment order by $[\alpha] \vdash [\beta]$ when $\alpha \vdash \beta$, and define the quotient orthogonality relation by $[\beta] \perp [\alpha]$ when $\beta \perp \alpha$ is provable. Finally, define the quotient tensor negation by $\neg[\alpha] \stackrel{\text{df}}{=} [\neg\alpha]$. These operations and relations are well-defined, and the tensors are associative. Since term tensor product and sum are monotonic w.r.t. entailment order, the tensor product and sum of quotient terms are also monotonic w.r.t. entailment order. So there is a join bisemilattice $\mathbf{Term}^{\oplus}_{\otimes} = \langle\langle \mathbf{Term}, \vdash, \otimes, \mathrm{Id}\rangle, \oplus, 0\rangle$ called the *quotient term pole*, whose objects are (quotients of) types, whose arrows are quotient terms, whose composition is the tensor product of quotients, and whose homset order is quotient entailment. Similarly, there is a meet bisemilattice $\mathbf{Term}^{\&}_{\nabla} = \langle\langle \mathbf{Term}, \vdash, \nabla, \mathrm{Id}\rangle, \&, 1\rangle$ called the *quotient term antipole*. Tensor negation is an involution of join bisemilattices $\mathbf{Term}^{\oplus}_{\otimes} \stackrel{\neg}{\rightharpoonup} (\mathbf{Term}^{\&}_{\nabla})^{\mathrm{coop}}$, and also an involution of meet bisemilattices $(\mathbf{Term}^{\oplus}_{\otimes})^{\mathrm{coop}} \stackrel{\neg}{\leftharpoonup} \mathbf{Term}^{\&}_{\nabla}$. So the two quotient term poles and quotient tensor negation form a polarized bisemilattice, also denoted by **Term**, for which the orthogonality-entailment axiom and the orthogonality definition axiom hold.

THEOREM 2. *The category* **Term** *of quotient terms is a Boolean category.*

The DeMorgan's law $\neg(\otimes\alpha) = \nabla(\vec{\neg}\alpha)$ states that the pair of *tensor term* operations is a morphism of polarized bipreorders $\mathbf{Lang}^* \stackrel{\langle\otimes,\nabla\rangle}{\longrightarrow} \mathbf{Term}$. It is a quotient functor (a full functor which is a bijection on objects), which constructs **Term** as the entailment-quotient category of $\mathbf{Lang}^*$.

Soundness and completeness. A *classical structure* $\langle \mathfrak{J}, \mathbf{B}\rangle$ for the basic calculus, the internal language of classical dialectical logic, consists of a Boolean category **B** and an interpretation map (graph morphism) $\mathbf{Lang} \stackrel{\mathfrak{J}}{\longrightarrow} \mathbf{B}$ which preserves negation, identities, entailment order, zeroes, ones, Boolean products and sums, and tensor products and sums. The interpretation map $\mathfrak{J}$ assigns to each type symbol x a **B**-type $\mathfrak{J}(x)$ and assigns to each atom $y \stackrel{a}{\rightharpoonup} x$ a **B**-term $\mathfrak{J}(y) \stackrel{\mathfrak{J}(a)}{\rightharpoonup} \mathfrak{J}(x)$. Following the polar sense of interpretation, we extend the interpretation $\mathfrak{J}$ to sequents by defining $\mathfrak{J}_{\otimes}(\alpha) \stackrel{\text{df}}{=} \mathfrak{J}(\otimes\alpha)$ for any sequent $y \stackrel{\alpha}{\rightharpoonup} x$. So $\mathfrak{J}$ is a morphism of polarized bipreorders $\mathbf{Lang}^* \stackrel{\mathfrak{J}}{\longrightarrow} \mathbf{B}$, with the polar interpretation embodied in the polar part $\mathbf{Lang}^* \stackrel{\mathfrak{J}_{\otimes}}{\longrightarrow} \mathbf{B}^{\oplus}_{\otimes}$ of $\mathfrak{J}$ (a morphism of bipreorders), and the antipolar interpretation embodied in the antipolar part $\mathbf{Lang}^* \stackrel{\mathfrak{J}_{\nabla}}{\longrightarrow} \mathbf{B}^{\&}_{\nabla}$ of $\mathfrak{J}$ (which is defined by $\mathfrak{J}_{\nabla} \stackrel{\text{df}}{=} \neg \cdot (\mathfrak{J}_{\otimes})^{\mathrm{coop}} \cdot (\neg_B)^{\mathrm{coop}}$). $\mathfrak{J}_{\otimes}$ preserves order, since if $\beta \vdash \alpha$ for any two parallel sequents $y \stackrel{\beta,\alpha}{\rightharpoonup} x$ then $\mathfrak{J}_{\otimes}(\beta) = \mathfrak{J}(\otimes\beta) \preceq \mathfrak{J}(\otimes\alpha) = \mathfrak{J}_{\otimes}(\alpha)$. Since $\beta \vdash\!\dashv \alpha$ implies $\mathfrak{J}_{\otimes}(\beta) = \mathfrak{J}_{\otimes}(\alpha)$ for any two parallel sequents $y \stackrel{\beta,\alpha}{\rightharpoonup} x$, there is a functor

$\mathbf{Term}^{\oplus}_{\otimes} \xrightarrow{\mathfrak{J}^{\oplus}_{\otimes}} \mathbf{B}^{\oplus}_{\otimes}$ uniquely satisfying the functorial equation $\mathfrak{J}_{\otimes} = \otimes(\ ) \cdot \mathfrak{J}^{\oplus}_{\otimes}$. The extended interpretation $\mathfrak{J}^{\oplus}_{\otimes}$ is the polar part of a morphism of Boolean categories $\mathbf{Term} \xrightarrow{\mathfrak{J}} \mathbf{B}$. The antipolar part, using the antipolar interpretation and tensor sum terms, is defined by $\mathfrak{J}^{\triangle}_{\nabla} \stackrel{\mathrm{df}}{=} \neg \cdot (\mathfrak{J}^{\oplus}_{\otimes})^{\mathrm{coop}} \cdot (\neg_B)^{\mathrm{coop}}$. The entailment quotient and the term category define the fundamental classical structure $\langle [\ ], \mathbf{Term} \rangle$, whose extended interpretation is the identity functor $[\ ]^{\oplus}_{\otimes} = \mathrm{Id}_{\mathrm{Term}}$.

THEOREM 3. *The Boolean category* **Term** *is free (w.r.t. the connectives) over the language (type-atom graph)* **Lang**.

An orthogonality assertion $\beta \perp \alpha$, for two opposed sequents $y \xrightarrow{\alpha} x$ versus $y \xleftarrow{\beta} x$, is (*tensorially*) *valid* in a structure $\mathfrak{J}$ when the orthogonality $\mathfrak{J}(\beta) \perp \mathfrak{J}(\alpha)$ holds in the Boolean category **B**. As a special case, an endosequent $x \xrightarrow{\alpha} x$ is valid in $\mathfrak{J}$ when $\mathfrak{J}(\alpha) \preceq \mathfrak{J}(x)$. A *tautology* is an orthogonality assertion $\beta \perp \alpha$ which is valid in any classical structure.

THEOREM 4 (Soundness). *The basic calculus for dialectical logic is sound w.r.t. validity in classical structures.*

THEOREM 5 (Completeness). *The basic calculus for dialectical logic is complete w.r.t. validity in classical structures.*

PROOF. Suppose $\beta \perp \alpha$ is a tautology at x. Then, since $\beta \perp \alpha$ is valid in every classical structure, it is valid in the free classical structure $\langle [\ ], \mathbf{Term} \rangle$, and so the orthogonality $[\beta] \perp [\alpha]$ holds in **Term**. But by definition, $[\beta] \perp [\alpha]$ iff $\beta \perp \alpha$ is provable. □

Summary. In this paper we have discussed the internal process aspect of dialectical logic, which is the logic of the flow dialectic. In the promised extension [Kent88] of this paper we will also discuss the external object aspect of dialectical logic, which is the logic of the flow constraint dialectic. This external aspect involves the semantic notions of monoids (preorder objects), processes, topologies and topomonoidal structures, and the axiomatic notions of exponentials (Girard's affirmation and consideration modalities) and quantifiers.

A. Subtypes

Comonoids. For any type x in a bisemilattice **P** a *comonoid* u at x, denoted by $u : x$, is an endoterm $x \xrightarrow{u} x$ which satisfies the "part" axiom (coreflexivity) $u \preceq_{x,x} x$, stating that u is a part of the type (identity term) x, and the "idempotency" axiom (cotransitivity) $u \preceq_{x,x} u \circ u$. A comonoid is also called an *interior term*. Since $u \circ u \preceq x \circ u = u$, we can replace the

inequality in the idempotency axiom with the equality $u \circ u = u$. For a functional term (adjoint pair) $y \overset{f \dashv f^{op}}{\rightarrow} x$ the composite interior endoterm $x \overset{f^{op} \circ f}{\rightarrow} x$ is called the *comonoid* of the functional term f. This comonoid is the top comonoid $f^{op} \circ f = x$ iff f is an epimorphism iff $f \dashv f^{op}$ is a reflective pair. The comonoids $y \overset{poi}{\rightarrow} y$ of subtypes $y \overset{i \dashv p}{\rightarrow} x$ are special x-comonoids which split (through y). In this sense comonoids are generalized subtypes. Comonoids of type x are ordered by entailment $\preceq_x \overset{df}{=} \preceq_{x,x}$. The bottom endoterm $\bot_x$ is the smallest comonoid of type x. The join $v \vee u$ of any two comonoids v, u of type x is also a comonoid of type x. Denote the join semilattice of comonoids of type x by $\Omega(x)$. We can interpret the semilattice $\Omega(x)$ as a "state-set" indexed by the type x, with a comonoid $u \in \Omega(x)$ being a "state" of a system. The state $u \in \Omega(x)$ has internal structure and is a composite object sharing an ordering of nondeterminism $\preceq_x$ with other states.

For any two comonoids $u, v \in \Omega(x)$ the tensor product is a lower bound $u \circ v \preceq u$ and $u \circ v \preceq v$ which is an upper bound for comonoids below u and v: if $w \preceq u$ and $w \preceq v$ then $w \preceq u \circ v$. If u and v commute $u \circ v = v \circ u$ then the tensor product $u \circ v$ is a comonoid; in which case it is the meet $u \circ v = u \wedge v$ in $\Omega(x)$. **[Standardization property:]** the bisemilattice **P** is said to be *locally standard* when $\Omega(x)$ is closed under tensor product for each type x; that is, when the tensor product $u \circ v$ is a comonoid for any two comonoids $u, v \in \Omega(x)$. Then $\Omega(x)$ is a lattice, with the tensor product $v \circ u$ of two comonoids $v, u \in \Omega(x)$ being the lattice meet in $\Omega(x)$, and the tensor product identity (or type) endoterm x being the largest comonoid of type x. Furthermore, the meet distributes over the join. We assume that any join bisemilattice **P** is locally standard. This standardization property means that the local contexts (monoidal semilattices) of comonoids $\{\Omega(x) \mid x$ a type$\}$ are standard contexts (distributive lattices).

In a complete Heyting category **H** an endoterm $x \overset{p}{\rightarrow} x$ contains a largest comonoid of the same type x, called the *interior* of p and denoted by p°. The interior is defined as the join $p^\circ \overset{df}{=} \bigvee \{w \in \Omega(x) \mid w \preceq_x p\}$, and satisfies the condition $w \preceq_x p$ iff $w \preceq_x p^\circ$ for all comonoids $w \in \Omega(x)$. In an arbitrary join bisemilattice **P**, we use this condition to define (and to assert the existence of) the interior of endoterms. The interior p°, when it exists, is the largest generalized **P**-subtype inside p. The interior of endoterms models the "affirmation modality" of linear logic [Girard]. Any comonoid $w \in \Omega(x)$ is its own interior $w^\circ = w$. Without the local standardization assumption, meets would still exist in $\Omega(x)$: the interior of the tensor product is the meet $(u \circ v)^\circ = u \wedge v = (v \circ u)^\circ$.

We are especially interested in join bisemilattices **P** for which any **P**-endoterm has such an interior. Such bisemilattices can be called interior (or affirmation) bisemilattices. A join bisemilattice **P** is an *interior bisemilattice* when at each type x the inclusion-of-comonoids monotonic function

$\Omega(x) \xrightarrow{\mathrm{Inc}_x} \mathbf{P}[x,x]$ has a right adjoint $\mathbf{P}[x,x] \xrightarrow{(\ )^\circ} \Omega(x)$ called *interior*, which with inclusion forms a coreflective pair of monotonic functions $\mathrm{Inc}_x \dashv (\ )^\circ$. Composition $(\ )^\circ \cdot \mathrm{Inc}_x$ is a general interior operator on endoterms. Any meets that exist in $\mathbf{P}[x,x]$ are preserved by interior $(p \wedge q)^\circ = p^\circ \circ q^\circ$ for endoterms $p, q \in \mathbf{P}[x,x]$, since interior is a right adjoint. In an interior Heyting category $\mathbf{H}$, the distributive lattice of comonoids $\Omega(x)$ at each type x is actually a complete Cartesian Heyting monoid, which is another name for a complete Heyting algebra. Since interiors exist, for any two comonoids $u, v \in \Omega(x)$ we can make the definition $u \Rightarrow v \stackrel{\mathrm{df}}{=} (u \backslash v)^\circ$. Then $u \Rightarrow v = (u \backslash v)^\circ = (v / u)^\circ$ is a locally standard implication, since $w \preceq u \Rightarrow v$ iff $w \preceq (u \backslash v)^\circ$ iff $w \preceq (u \backslash v)$ iff $u \circ w \preceq v$ iff $w \circ u \preceq v$ iff $w \preceq (v / u)$ iff $w \preceq (v / u)^\circ$. Comonoids in bisemilattices, and even more strongly in interior Heyting categories, play the role of "localized truth values". Any complete Heyting category is an interior Heyting category.

In a bisemilattice $\mathbf{P}$, for each $\mathbf{P}$-adjunction (functional term) $y \xrightarrow{f \dashv f^{\mathrm{op}}} x$ and each $\mathbf{P}$-comonoid $v \in \Omega(y)$ at y, the endoterm $x \xrightarrow{f^{\mathrm{op}} \circ v \circ f} x$ is a $\mathbf{P}$-comonoid $(f^{\mathrm{op}} \circ v \circ f) \in \Omega(x)$ at x. So the direct image monotonic function $\mathbf{P}^f$ restricts to $\mathbf{P}$-comonoids. Denote this restriction by $\Omega(y) \xrightarrow{\Omega^f} \Omega(x)$ and call it the *direct image* also. When $\mathbf{P}$ is an interior bisemilattice, the direct image function has a right adjoint $\Omega(y) \xleftarrow{\Omega_f} \Omega(x)$ called the *inverse image* monotonic function, and defined by $\Omega_f(u) \stackrel{\mathrm{df}}{=} (f \circ u \circ f^{\mathrm{op}})^\circ$ for each $\mathbf{P}$-comonoid $u \in \Omega(x)$. If we denote this adjointness by $\Omega(f) \stackrel{\mathrm{df}}{=} (\Omega^f \dashv \Omega_f)$, then the comonoid construction Ω is an indexed adjointness (dialectical base) $\mathbf{P}^{\dashv} \xrightarrow{\Omega} \mathbf{adj}$, mapping functional $\mathbf{P}$-terms into the subcategory of $\mathbf{adj}$ consisting of distributive lattices and adjoint pairs of monotonic functions.

In subset categories $\mathcal{P}(\mathbf{C})$ a comonoid of type x is either the empty endoterm $x \xrightarrow{\emptyset} x$ or the identity singleton $x \xrightarrow{\{x\}} x$, and these can be interpreted as the truth-values **false** and **true**, so that $\Omega(x)$ is the complete Heyting algebra $\Omega(x) \cong \mathbf{2}$. In closure subset categories $\mathcal{P}(\mathbf{P})$ a comonoid $x \xrightarrow{W} x$ of type x is a closed-below subset $W \subseteq \mathbf{P}[x,x]$ of $\mathbf{P}$-endoterms $x \xrightarrow{w} x$, which are subparts of the identity $w \preceq x$ and which factor (possibly trivially) $w \preceq v \circ u$ into two other endoterms $v, u \in W$. Since $\mathcal{P}(\mathbf{P})$ is a cHc, the lattice of comonoids $\Omega_{\mathcal{P}(\mathbf{P})}(x)$ is also a complete Heyting algebra. Any $\mathbf{P}$-comonoid $x \xrightarrow{w} x$ is embeddable as the $\mathcal{P}(\mathbf{P})$-comonoid $x \xrightarrow{\downarrow w} x$. So we can regard $\mathcal{P}(\mathbf{P})$-comonoids as generalized $\mathbf{P}$-comonoids called *closure subset* $\mathbf{P}$*-comonoids.*

For any source and target comonoids $v \in \Omega(y)$ and $u \in \Omega(x)$ the term $v \xrightarrow{r_{vu}} u$ defined by $r_{vu} \stackrel{\mathrm{df}}{=} v \circ r \circ u$ is called the (v,u)*-th subterm* of r. A $\mathbf{P}$*-coprocess* $v \xrightarrow{r} u$ is a $\mathbf{P}$-term $y \xrightarrow{r} x$ which satisfies the external source constraint $v \circ r \succeq_{y,x} r$ saying that r restricts to the source comonoid $v : y$, and which

satisfies the external target constraint $r \circ u \succeq_{y,x} r$ saying that r corestricts to the target comonoid $u : x$. The source/target restriction conditions can be replaced by the two equalities $v \circ r = r$ and $r \circ u = r$; or by the single equality $r_{vu} = v \circ r \circ u = r$. Thus, the notion of coprocess allows comonoids to function as identity arrows, or objects, of some category. To make this precise we define the biposet $\Omega(\mathbf{P})$, whose objects are $\mathbf{P}$-comonoids and whose arrows are $\mathbf{P}$-coprocesses. Although $\Omega(x) \subseteqq \mathbf{P}[x,x]$, note that $\Omega(x) \neq \mathbf{P}[x,x]$, since endoarrows exist which are not comonoids. Given any $\mathbf{P}$-term $y \xrightarrow{r} x$, let $\mathcal{F}_0(r) \subseteqq \Omega(y)$ denote the collection $\mathcal{F}_0(r) \stackrel{\text{df}}{=} \{v \mid v \circ r \succeq_{y,x} r\}$ of all comonoids at the source type y satisfying source restriction. Since $\mathcal{F}_0(r)$ is closed above and closed under finite meets (= tensor products) it is a filter in the lattice $\Omega(y)$ called the *source filter* of r. Similarly, the *target filter* $\mathcal{F}_1(r)$ of r is the collection $\mathcal{F}_1(r) \stackrel{\text{df}}{=} \{u \mid r \preceq_{y,x} r \circ u\} \subseteqq \Omega(x)$ of all comonoids at x satisfying target corestriction. Given two comonoids $v : y$ and $u : x$, a term $y \xrightarrow{r} x$ is a coprocess $v \xrightarrow{r} u$ iff $v \in \mathcal{F}_0(r)$ and $u \in \mathcal{F}_1(r)$.

Unfortunately, the category $\Omega(\mathbf{P})$ is not as useful as one might desire; in particular, there is no canonical functor to the underlying category $\mathbf{P}$ of types and terms since identities are not preserved. But by suitably weakening the constraint $v \circ r = r = r \circ u$ we get a very useful and interesting category. A *Hoare triple* or *Hoare assertion* $v : y \xrightarrow{r} u : x$, denoted traditionally although imprecisely by $\{v\}r\{u\}$, consists of a "flow specifying" $\mathbf{P}$-term $y \xrightarrow{r} x$ and two $\mathbf{P}$-comonoids, a "precondition" or source comonoid $v \in \Omega(y)$ and a "postcondition" or target comonoid $u \in \Omega(x)$, which satisfy the "precondition/postcondition constraint" $v \circ r \preceq r \circ u$. Clearly, composition of Hoare triples $\{w\}s\{v\} \circ \{v\}r\{u\} = \{w\}(s \circ r)\{u\}$ is well-defined and $\{u\}x\{u\}$ is the identity Hoare triple at the comonoid $u : x$. Also, there is a zero triple $\{v\}0_{y,x}\{u\}$ for any precondition $v \in \Omega(y)$ and postcondition $u \in \Omega(x)$, and if $\{v\}r\{u\}$ and $\{v\}s\{u\}$ are two triples with the same precondition and postcondition then $\{v\}(r \oplus s)\{u\}$ is also a triple. So typed comonoids as objects and Hoare triples as arrows form a join bisemilattice $\mathcal{H}(\mathbf{P})$ called the *Hoare assertional category* over $\mathbf{P}$. There is an obvious underlying type/term functor $\mathcal{H}(\mathbf{P}) \xrightarrow{T_P} \mathbf{P}$ which is a morphism of join bisemilattices. For each type x in $\mathbf{P}$, the *fiber* over x is the subcategory $T_P^{-1}(x) \subseteqq \mathcal{H}(\mathbf{P})$ of all comonoids and triples which map to x. The objects in $T_P^{-1}(x)$ are the comonoids of type x and the triples in $T_P^{-1}(x)$ are of the form $\{u'\}x\{u\}$, pairs of comonoids of type x satisfying $u' \preceq u$. Hence, the fiber over x is just the join semilattice (actually, lattice) of comonoids $T_P^{-1}(x) = \Omega(x)$. The axiomatics, semantics and dialectics of Hoare assertional categories and associated constructions, and their relationship to dynamic logic, is explored in detail in [Kent89].

Topotypes and topomatrices. The closure subset construction $\mathcal{P}(\mathbf{P})$ does not capture the notion of "relational structures" completely. Although it introduces nondeterminism on the arrows, it leaves the objects alone. The

notions of "topology" and "subtype" can be naturally combined and locally defined in any cHc $\mathbf{H}$. Topologies of subtypes introduce distributivity on objects. A *topology of* $\mathbf{H}$*-comonoids* or $\mathbf{H}$*-topotype* $W = \langle W, x\rangle$, denoted by $W : x$, is a topology W in the complete lattice $\Omega(x)$ of comonoids at x regarded as a one-object subcategory of $\mathbf{H}$ (the more general notion of a *topology* in a cHc $\mathbf{H}$ is discussed in [Kent88]); that is, W is a collection $W \subseteq \Omega(x)$ of comonoids of x, which is closed under finite tensor products and arbitrary homset joins. A topotype is a kind of "power type", which is *not* imposed from without, but arises naturally out of the mathematical structure. Since tensor products are finite homset meets for comonoids, a topotype $W : x$ is just a standard topology in the complete lattice $\Omega(x)$. An advantage of standard topologies over general tensor product topologies is that homset order is more directly related to topological meet. W is interpreted to be an *object of inner truth-values* at type x, and its topological nature can be used to define approximation or limit structures on terms whose source or target is x. Any comonoid $u : x$ can be identified with the topotype $u = \{\bot_x, u, x\}$.

A topomatrix is a matrix indexed by topologies. Given two topotypes $V : y$ and $U : x$, an $\mathbf{H}$*-topomatrix* $V : y \xrightarrow{R} U : x$, denoted by $R = (r_{vu} \mid v \in V, u \in U)$, is an $\Omega(\mathbf{H})$-matrix $V \times U \xrightarrow{R} \mathrm{Ar}(\Omega(\mathbf{H}))$ monotonically indexed by the source and target topologies. Monotonic indexing means that if $v \preceq v'$ and $u \preceq u'$ then $r_{vu} \preceq r_{v'u'}$. This monotonic indexing property is similar to the compatibility of ordinary partial functions on the overlap of their domains of definition. Every cHc $\mathbf{H}$ has an associated *category of topomatrices* $\mathcal{M}_{\mathcal{T}}(\mathbf{H})$, whose objects are topotypes $U : x$, whose arrows $V : y \xrightarrow{R} U : x$ are topomatrices, whose homset order is pointwise order $(s_{vu}) \preceq (r_{vu})$ when $s_{vu} \preceq r_{vu}$ for all $v \in V$ and $u \in U$, whose tensor product is the matrix product $(S \circ R)_{wu} \stackrel{\mathrm{df}}{=} \bigvee_{v \in V}[s_{wv} \circ r_{vu}]$, and whose identity at $U : x$ is the topomatrix $(u' \circ u = u' \wedge u \mid u', u \in U)$. The join operator is a *join* functor $\mathcal{M}_{\mathcal{T}}(\mathbf{H}) \xrightarrow{\vee} \mathbf{H}$, which maps each topotype to its underlying type $\bigvee(U : x) = x$ and maps each $V \times U$ topomatrix $R = (r_{vu})$ to its *join term* $\bigvee R = \bigvee_{v \in V, u \in U} r_{vu}$, the join of all the coprocess entries in R. The (V, U)-th component of the join functor $\bigvee$ is a *join* join-continuous monotonic function $\mathcal{M}_{\mathcal{T}}(\mathbf{H})[V : y, U : x] \xrightarrow{\vee_{V,U}} \mathbf{H}[y, x]$. The category of comonoids $\Omega(\mathbf{H})$ can be embedded $\Omega(\mathbf{H}) \xrightarrow{\mathrm{Inc}} \mathcal{M}_{\mathcal{T}}(\mathbf{H})$ into the category of topomatrices $\mathcal{M}_{\mathcal{T}}(\mathbf{H})$ by $\mathrm{Inc}(u : x) = \{\bot, u, x\} : x$ and $\mathrm{Inc}(v : y \xrightarrow{r} u : x) = \{(\bot, \bot, \bot), (\bot, \bot, u), (\bot, \bot, x), (v, \bot, \bot), (y, \bot, \bot)\} \cup \{(v, r, u)\} \cup \cup \{(v, r, x), (y, r, u), (y, r, x)\}$. The composition of comonoid embedding with join is the underlying type functor $\mathrm{Inc} \cdot \bigvee = U_H$. The restriction of the comonoid-as-topology embedding to identity comonoids defines the *indiscrete-topology functor* $\mathbf{H} \xrightarrow{\{\}} \mathcal{M}_{\mathcal{T}}(\mathbf{H})$, *where* $\{x\} = \{\bot, x\} : x$ *and* $\{r\} = \{(\bot, \bot, \bot), (\bot, \bot, x), (y, \bot, \bot)\} \cup \{(y, r, x)\}$. This functor is clearly ful-

ly-faithful, since for two fixed types y and x, there is a bijection $\mathbf{H}[y,x] \cong \mathcal{M}_{\mathcal{T}}(\mathbf{H})[\{y\},\{x\}]$. Also, $\{\} \cdot \bigvee = \mathrm{Id}_H$. This implies that the join functor is surjective on objects.

A representation theorem. Let $V : y$ and $U : x$ be any two $\mathbf{H}$-topotypes, and let $y \xrightarrow{r} x$ be any $\mathbf{H}$-term. The topomatrix $V : y \xrightarrow{(r)_V^U} U : x$ defined by $(r)_V^U \stackrel{\mathrm{df}}{=} \left(v \xrightarrow{r_{vu}} u \mid v \in V, u \in U\right)$, where $r_{vu} \stackrel{\mathrm{df}}{=} v \circ r \circ u$ is the (v,u)-th subterm of r, is called the *decomposition matrix* of r. Such decompositions, especially w.r.t. topological bases of comonoids, give an internal representation of cHc's as distributor-like categories. This defines a *decomposition* join-continuous monotonic function $\mathbf{H}[y,x] \xrightarrow{\#_{V,U}} \mathcal{M}_{\mathcal{T}}(\mathbf{H})[V : y, U : x]$, where $\#_{V,U}(r) \stackrel{\mathrm{df}}{=} (r)_V^U$. Moreover, any $\mathbf{H}$-term $y \xrightarrow{r} x$ is recoverable from its decomposition matrix $(r)_V^U$ by applying the join functor $\bigvee_{V,U}(\#_{V,U}(r)) = \bigvee_{V,U}((r)_V^U) = \bigvee_{v \in V, u \in U} r_{v,u} = r$. This means that the join functor is full (surjective on arrows). Conversely, an $\mathbf{H}$-topomatrix $V : y \xrightarrow{R} U : x$ is recoverable from its join term $\bigvee R$ by applying the partition function $\#_{V,U}(\bigvee_{V,U}(R)) = R$. This means that the join functor is faithful (injective on arrows). So for two fixed topotypes $V : y$ and $U : x$, the decomposition and join monotonic functions are inverse to each other, and define an isomorphism $\mathbf{H}[y,x] \cong \mathcal{M}_{\mathcal{T}}(\mathbf{H})[V : y, U : x]$.

LEMMA 2. *The join functor* $\mathcal{M}_{\mathcal{T}}(\mathbf{H}) \xrightarrow{\vee} \mathbf{H}$ *is fully-faithful, and a surjection on objects.*

A topomatrix $V : y \xrightarrow{R} \{x\}$ is called a *column* $\mathbf{H}$*-topovector*. If $y \xrightarrow{r} x$ is any term and $V : y$ is a topology at y, then the V*-source decomposition* of r is the column topovector $V : y \xrightarrow{]r[_V} \{x\}$ defined by $]r[_V \stackrel{\mathrm{df}}{=} \left(v \xrightarrow{r_{vx}} x \mid r_{vx} = v \circ r, v \in V\right)$. The V*-source cotupling* of a column topovector $V : y \xrightarrow{R} \{x\}$, where R is the V-indexed collection of coprocesses $\left(v \xrightarrow{r_{vx}} x \mid v \in V\right)$, is the $\mathbf{H}$-term $y \xrightarrow{[R]_V} x$ defined by $[R]_V \stackrel{\mathrm{df}}{=} \bigvee_{v \in V} r_{vx}$. The source decomposition and cotupling operations are inverse to each other, with $[\,]r[V]_V = r$ and $][R]_V[_V = R$. Dually, a topomatrix $\{y\} \xrightarrow{R} U : x$ is called a *row* $\mathbf{H}$*-topovector*. If $y \xrightarrow{r} x$ is any term and $U : x$ is a topology at x, then the U*-target decomposition* of r is the row topovector $\{y\} \xrightarrow{\rangle r\langle^U} U : x$ defined by $\rangle r\langle^U \stackrel{\mathrm{df}}{=} \left(y \xrightarrow{r_{yu}} u \mid r_{yu} = r \circ u, u \in U\right)$. The U*-target tupling* of a row topovector $\{y\} \xrightarrow{R} U : x$, where R is the U-indexed collection of coprocesses $\left(y \xrightarrow{r_{yu}} u \mid u \in U\right)$, is the $\mathbf{H}$-term $y \xrightarrow{\langle R\rangle^U} x$ defined by $\langle R\rangle^U \stackrel{\mathrm{df}}{=} \bigvee_{u \in U} r_{yu}$. The target decomposition and tupling

operations are inverse to each other, with $\langle\rangle r\langle^U\rangle^U = r$ and $\rangle\langle R\rangle^U\langle^U = R$.

Any topology $U : x$ at x decomposes the identity term $x \xrightarrow{x} x$ in either of two ways: as the source decomposition column topovector $U : x \xrightarrow{\iota_U} \{x\}$ defined by $\iota_U \overset{\text{df}}{=}]x[_U = \left(u \xrightarrow{u} x \mid u \in U\right)$, or as the target decomposition row topovector $\{x\} \xrightarrow{\pi_U} U : x$ defined by $\pi_U \overset{\text{df}}{=} \rangle x\langle^U = \left(x \xrightarrow{u} u \mid u \in U\right)$. Moreover, the identity matrix at $U : x$ decomposes as $\iota_U \circ \pi_U$, and the identity matrix at $\{x\}$ decomposes as $\pi_U \circ \iota_U$, so that $U : x \xrightarrow{\iota_U} \{x\}$ and $\{x\} \xrightarrow{\pi_U} U : x$ are inverse topomatrices. Since ι_U and π_U are inverse pairs, they are adjoint pairs in both directions $U : x \overset{\iota_U \dashv \pi_U}{\longrightarrow} \{x\}$ and $\{x\} \overset{\pi_U \dashv \iota_U}{\longrightarrow} U : x$. So, given any term $y \xrightarrow{r} x$ and any topotypes $V : y$ and $U : x$, (1) the term r and its source decomposition $]r[_V$ are expressible in terms of each other via the direct and inverse left flow expressions $]r[_V = \iota_V \circ \{r\} = \pi_V \backslash \{r\}$ and $\{r\} = \pi_V \circ]r[_V = \iota_V \backslash]r[_V$, and (2) the term r and its target decomposition $\rangle r\langle^U$ are expressible in terms of each other via the direct and inverse right flow expressions $\rangle r\langle^U = \{r\} \circ$ $\circ \pi_U = \{r\} / \iota_U$ and $\{r\} = \rangle r\langle^U \circ \iota_U = \rangle r\langle^U / \pi_U$. Furthermore, given any two topotypes $V : y$ and $U : x$, (1) a term $y \xrightarrow{r} x$ and its decomposition matrix $\#_{V,U}(r) = (r)^U_V$ are expressible in terms of each other via the direct flow expressions $r = \pi_V \circ \#_{V,U}(r) \circ \iota_U$ and $\#_{V,U}(r) = \iota_V \circ \{r\} \circ \pi_U$, and (2) an **H**-topomatrix $V : y \xrightarrow{R} U : x$ and its join term $y \xrightarrow{\vee R} x$ are expressible in terms of each other via the direct flow expressions $R = \iota_V \circ \{\bigvee R\} \circ \pi_U$ and $\bigvee R =$ $= \pi_V \circ R \circ \iota_U$.

For each topotype $U : x$ the topomatrix isomorphism $\{x\} \xrightarrow{\pi_U} U : x$ is the $(U : x)$-th component of a "counit" natural isomorphism $\pi : \bigvee \cdot \{\} \Longrightarrow$ $\Rightarrow \mathrm{Id}_{\mathcal{M}_T(H)}$, since $\{\bigvee R\} \circ \pi_U = \pi_V \circ R$.

THEOREM 6. *For every cHc* **H**, *the indiscrete-topology and join functors form a categorical equivalence* $\{\} \dashv \bigvee$ *between* **H** *and its category of topomatrices* $\mathcal{M}_T(\mathbf{H})$, *with identity unit* $\mathrm{Id}_H = \{\} \cdot \bigvee$ *and natural isomorphism counit* $\pi : \bigvee \cdot \{\} \Longrightarrow \mathrm{Id}_{\mathcal{M}_T(H)}$.

Given three topotypes $W : z$, $V : y$ and $U : x$ and two terms $z \xrightarrow{s} y$ and $y \xrightarrow{r} x$, the (w, u)-th subterm $(s \circ r)_{wu}$ is the join $(s \circ r)_{wu} = \bigvee_{v \in V} s_{wv} \circ r_{vu}$, so that decomposition maps tensor products of terms to products of matrices $(s)^V_W \circ (r)^U_V = (s \circ r)^U_W$. Also, the $U \times U$ decomposition matrix of the identity term $x \xrightarrow{x} x$ is the identity matrix $(x)^U_U = \iota_U \circ \pi_U$, where $(x)^U_{U_{u'u}} = u' \circ u =$ $= u' \wedge u$. The type x is a direct sum of V-open comonoids when $x = \bigvee X$ for some collection $X \subseteqq V$ of pairwise disjoint comonoids.

Let **W** be a standard topology on the lattice of all **H**-comonoids $\Omega(\mathbf{H})$. **W** can be partitioned into a collection of topotypes $\mathbf{W} = \{\mathbf{W}(x) \subseteqq \Omega(x) \mid x \in$ $\in \mathrm{Obj}(\mathbf{H})\}$. We call such a collection **W** a *topotypeal structure.* A topotypeal structure is a "choice functor", choosing a topology at each **H**-type. Topo-

typeal structures are a type-indexed version of Girard's topolinear spaces in linear logic. Any topotypeal structure $\mathbf{W}$ defines, and can be identified with, an embedding $\mathbf{H} \xrightarrow{\#_W} \mathcal{M}_T(\mathbf{H})$, of $\mathbf{H}$ into its category of topomatrices $\mathcal{M}_T(\mathbf{H})$ called the $\mathbf{W}$-*decomposition* of terms. On types $\#_W = \mathbf{W}(x)$ is the x-th topotype of $\mathbf{W}$, and on terms $\#_W(r) = (r)^{W(x)}_{W(y)}$ is the $\mathbf{W}(y) \times \mathbf{W}(x)$ decomposition matrix of r. Partition followed by join is the identity functor $\#_W \cdot \vee = \mathrm{Id}_H$. The indiscrete-topology inclusion functor $\mathbf{H} \xrightarrow{\{\}} \mathcal{M}_T(\mathbf{H})$ is the decomposition functor $\{\} = \#_\Delta$ for the trivial topotypeal structure $\Delta = \{\{\bot, x\} \subseteqq \Omega(x) \mid x \in \mathrm{Obj}(\mathbf{H})\}$. For any topotypeal structure $\mathbf{W}$, the $\mathbf{W}$-*decomposition category* $\mathcal{M}_T(\mathbf{W}) \subseteqq \mathcal{M}_T(\mathbf{H})$, is the full subcategory which is the image of the $\mathbf{W}$-decomposition functor $\#_W$. There is a $\mathbf{W}$-*join functor* $\mathcal{M}_T(\mathbf{W}) \xrightarrow{\vee_W} \mathbf{H}$ which is the restriction of join $\bigvee$ to $\mathbf{W}$-matrices $\mathcal{M}_T(\mathbf{W})$, and a W-*decomposition functor* $\mathbf{H} \xrightarrow{\#_W} \mathcal{M}_T(\mathbf{W})$ which is the corestriction of $\mathbf{W}$-decomposition $\#_W$ to $\mathbf{W}$-matrices $\mathcal{M}_T(\mathbf{W})$. For a fixed topotypeal structure $\mathbf{W}$, these decomposition and join functors are inverse to each other.

THEOREM 7. *Any cHc* $\mathbf{H}$ *is isomorphic to each of its decomposition categories:* $\mathbf{H} \cong \mathcal{M}_T(\mathbf{W})$ *for any topotypeal structure* $\mathbf{W}$.

So each topotypeal structure $\mathbf{W}$ defines a representation of the cHc $\mathbf{H}$ inside of its category of topomatrices $\mathcal{M}_T(\mathbf{H})$; namely, $\mathcal{M}_T(\mathbf{W})$.

Flow decomposition. For any cHc $\mathbf{H}$, in the category of $\mathbf{H}$-topomatrices $\mathcal{M}_T(\mathbf{H})$ source and target tuplings are related to direct and inverse flow by the identities

$$\langle (t_{zv} \mid v \in V) \rangle^V \circ [(r_{vx} \mid v \in V)]_V = \bigvee_{v \in V} (t_{zv} \circ r_{vx} \mid v \in V)$$

"right tensor product along V-source tupling"

$$t \circ \langle (r_{yu} \mid u \in U) \rangle^U = \langle (t \circ r_{yu} \mid u \in U) \rangle^U$$

"right tensor product along U-target tupling"

$$[(r_{vx} \mid v \in V)]_V \circ s = [(r_{vx} \circ s \mid v \in V)]_v$$

"left tensor product along V-source tupling"

$$\langle (r_{yu} \mid u \in U) \rangle^U \circ [(s_{uz} \mid u \in U)]_U = \bigvee_{u \in U} (r_{yu} \circ s_{uz} \mid u \in U)$$

"left tensor product along U-target tupling"

$$s \backslash [(r_{vx} \mid v \in V)]_V = \langle (s \backslash r_{vx} \mid v \in V) \rangle^V$$

"right tensor implication along V-source tupling"

$$\langle (s_{zu} \mid u \in U) \rangle^U / \langle (r_{yu} \mid u \in U) \rangle^U = \bigwedge_{u \in U} (s_{zu} / r_{yu} \mid u \in U)$$

"right tensor implication along U-target tupling"

$$[(r_{vx} \mid v \in V)]_V \backslash \langle (t_{vz} \mid v \in V) \rangle^V = \bigwedge_{v \in V} (r_{vx} \backslash t_{vz} \mid v \in V)$$

"left tensor implication along V-source tupling"

$$\langle (r_{yu} \mid u \in U) \rangle^U \backslash t = [(r_{yu} \backslash t \mid u \in U)]_U$$

"left tensor implication along U-target tupling"

These identities reduce the action of direct and inverse term flow to components.

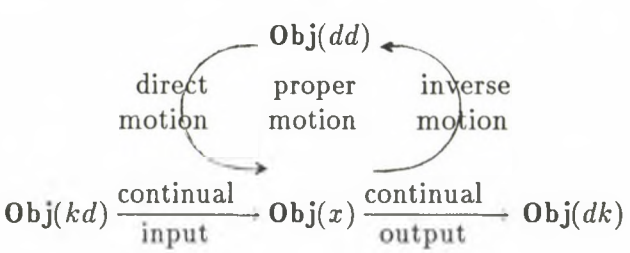
B. Dialectical reproduction

We work in a Heyting category **H**, and assume the existence of a special type 1 which is a *separator* of terms in the following sense: for any two parallel terms $y \xrightarrow{s,r} x$, if $\psi \circ s = \psi \circ r$ for all terms $1 \xrightarrow{\psi} y$ then $s = r$. A term $1 \xrightarrow{\phi} x$ is called an *object* of type x, and denoted by $\phi \in x$. In relational database theory, where the Heyting category **H** is the category of monoids and processes [Kent88] of closed subsets of Σ-terms, a monoid $m : x$ (**H**-type) represents a constrained database scheme consisting of database scheme x and semantic constraints m, and an m-object is a database which satisfies that scheme and those semantic constraints. In the general theory of dialectics, two possible meanings for "entities in dialectical motion" are (1) *comonoids* $u \in \Omega(x)$; and (2) *objects* $1 \xrightarrow{\phi} x$. Here we discuss the flow of objects in more detail. In a succeeding paper [Kent89] we will discuss the flow of comonoids, and we will also discuss the important notion of transformation between these two kinds of entities.

Let $\mathbf{Obj}(x)$ denote the lattice of all objects of type x with object order $\preceq_x \stackrel{\text{df}}{=} \preceq_{1,x}$; that is, $\mathbf{Obj}(x) = \mathbf{H}[1, x]$. Terms define a dialectical (bidirectional) flow of objects which is expressed in terms of tensor product and implication: for any term $y \xrightarrow{r} x$ let $\mathbf{Obj}^r = (\ ) \circ r$ denote right tensor product by r, and let $\mathbf{Obj}_r = (\ ) \backslash r$ denote right tensor implication by r. So $\mathbf{Obj}^r$ is the right direct flow and $\mathbf{Obj}_r$ is the right inverse flow of r. We identify this dialectical flow of objects as the *behavior* of the term r. The separator rule states that terms are distinguished (and can be identified) by their direct flow behavior. Direct flow $\mathbf{Obj}(y) \xrightarrow{\mathbf{Obj}^r} \mathbf{Obj}(x)$ and inverse flow $\mathbf{Obj}(y) \xleftarrow{\mathbf{Obj}_r} \mathbf{Obj}(x)$ are monotonic functions, and the dialectical axioms state that these form an adjoint pair $\mathbf{Obj}^r \dashv \mathbf{Obj}_r$. As noted before direct flow is "functorial", $\mathbf{Obj}^{s \circ r} = \mathbf{Obj}^s \cdot \mathbf{Obj}^r$ and $\mathbf{Obj}^x = \mathrm{Id}_{\mathbf{Obj}(x)}$, and inverse flow is "contravariantly functorial", $\mathbf{Obj}_{s \circ r} = \mathbf{Obj}_r \cdot \mathbf{Obj}_s$ and $\mathbf{Obj}_x = \mathrm{Id}_{\mathbf{Obj}(x)}$. In summary, if we combine the adjoint pairs as $\mathbf{Obj}(r) = (\mathbf{Obj}^r \dashv \mathbf{Obj}_r)$, then the above laws and rules are equivalent to the statement that the object concept or *flow dialectic* is functorial $\mathbf{H} \xrightarrow{\mathbf{Obj}} \mathbf{adj}$, mapping types to their object lattice and terms to their behavior. This is the sense in which terms specify the dialectical motion of objects.

So tensor product defines the *direct aspect* of term flow, whereas tensor implication defines the *inverse aspect.* As is clear now (manifested by the doubling of implication) and more clear latter (however, see Kelley's development of tensors using hom-objects), the direct aspect of flow is the principal aspect. This notion of principal aspect seems to occur often in applied dialectics. We develop here the full theory of dialectical terms. However, an interesting and coherent *direct subtheory* of terms, using only the direct aspect of flow, is included. This direct subtheory seems to include much of traditional process theory, but is impoverished by not having the concept of inverse flow.

Since the behavior of terms is identified with (dialectical) flow, either direct flow or inverse flow, one means of interaction/communication between terms is by flow composition. If we make the identification "types≡ports", then terms communicate through their source and/or target ports. A parallel pair of terms $y \overset{s,r}{\rightarrow} x$, a graph in a Heyting category, is known as a *dialectical system.* The dialectical interaction (complementary union) of the component terms of a dialectical system occurs through both source and target ports. The notion of *reproduction* in a system is specified by the dialectical flow (fixpoint operator) $\bigcirc_r^s(\ ) = ((\ )/\!\!-r) \circ s$. This reproduction operator can be interpreted as the "polar-turning structure" of the preSocratic Greek philosopher Heraclitus [Hussey], and in Greek is rendered $\pi\alpha\lambda\iota\nu\tau\rho o\pi o\varsigma\ \ \alpha\rho\mu o\nu\iota\eta$. An object ϕ is *reproduced* when it satisfies the fixpoint equation $\bigcirc_r^s(\phi) = \phi$. [A philosophical note: The notion of complementary union (two working together in one) is not that of "synthesis". Neither of the opposites is "transformed". Indeed, with synthesis, dialectical motion would cease! The notion of "reproduction" is one of equilibrium of motion, not lack of motion.] Here the yin-yang symbol $\bigcirc_r^s$ is used as a reminder of ancient dialectics; *yin* inverse flow along r and *yang* direct flow along s. Starting with (quotient) objects at the source type, there is an op-dual "reverse time" yin-yang fixpoint operator $(s \circ (r\backslash(\,)))$. There are also yang-yin operators with direct flow first and reverse flow last: To claim a type of uniqueness for reproduced objects ϕ we can use: the **least fixpoint rule** $\bigcirc_r^s(\phi) = \phi$, and if $\bigcirc_r^s(t) = t$ then $\phi \preceq t$; or the **greatest fixpoint rule** $\bigcirc_r^s(\phi) = \phi$, and if $\bigcirc_r^s(t) = t$ then $t \preceq \phi$. The system motion is graphically represented as follows:



where the collection of y-subtypes kd, dd, dk and kk consists of, respectively, the "atomic subtype", "proper subtype", "negative subtype" and "nil sub-

type" of the source type y. These correspond to clause types in Horn clause logic.

For any term $y \xrightarrow{r} x$, dialectical flow along r is decreasing: $\ominus_r^r(\phi) \preceq \phi$ for every object $1 \xrightarrow{\phi} x$. For any functional term $y \xrightarrow{f \dashv f^{\mathrm{op}}} x$, dialectical flow along f is equal to dialectical flow along the associated interior comonoid $f^{\mathrm{op}} \circ f$, $\ominus_f^f = \ominus_{f^{\mathrm{op}} \circ f}^{f^{\mathrm{op}} \circ f}$, since $(\ ) \circ f^{\mathrm{op}} = (\ ) \not\vdash f$ implies $\ominus_{f^{\mathrm{op}} \circ f}^{f^{\mathrm{op}} \circ f} = [(\ ) \not\vdash f] \cdot [(\ ) \not\vdash f^{\mathrm{op}}] \cdot [() \circ f^{\mathrm{op}}] \cdot [(\ ) \circ f] = [(\ ) \circ f^{\mathrm{op}}] \cdot [(\ ) \not\vdash f^{\mathrm{op}}] \cdot [() \circ f^{\mathrm{op}}] \cdot [(\ ) \circ f] \preceq (\succeq \succeq)[() \circ f^{\mathrm{op}}] \cdot [(\ ) \circ f] = [(\ ) \not\vdash f] \cdot [(\ ) \circ f] = \ominus_f^f$. This fact includes subtypes as a special case. So for dialectical flow along functional terms, we can restrict our attention to comonoids. Let $V : y$ be any topotype (topology of comonoids at y). The join of the dialectical flows of the topotype comonoids is unity $\bigvee_{v \in V} \ominus_v^v = \mathrm{Id}$, since $\psi = \psi \circ y = \psi \circ \left(\bigvee_{v \in V} v\right) = \bigvee_{v \in V} (\psi \circ v) = \bigvee_{v \in V} (\psi \not\vdash y) \circ v = \bigvee_{v \in V} \left(\psi \not\vdash \left(\bigvee_{v' \in V} v'\right)\right) \circ v = \bigvee_{v \in V} \left(\bigwedge_{v' \in V} (\psi \not\vdash v')\right) \circ v \preceq \preceq \bigvee_{v \in V} \bigwedge_{v' \in V} ((\psi \not\vdash v') \circ v) \preceq \bigvee_{v \in V} ((\psi \not\vdash v) \circ v) \preceq \bigvee_{v \in V} \psi = \psi$ for every y-object $1 \xrightarrow{\psi} y$.

FACT 2. *For any dialectical system $y \xrightarrow{s,r} x$ and any source topotype $V : y$, dialectical flow decomposes as*

$$\ominus_r^s = \bigvee_{v \in V} \ominus_{r_v}^{s_v} .$$

PROOF. $\bigvee_{v \in V} \ominus_{r_v}^{s_v} = \bigvee_{v \in V} [(\ ) \not\vdash (v \circ r)] \cdot [(\ ) \circ (v \circ s)] = \bigvee_{v \in V} [(\ ) \not\vdash r] \cdot [(\ ) \not\vdash v] \cdot [(\ ) \circ v] \cdot [(\ ) \circ s] = \bigvee_{v \in V} [(\ ) \not\vdash r] \cdot \ominus_v^v \cdot [() \circ s] = [(\ ) \not\vdash r] \cdot \left(\bigvee_{v \in V} \ominus_v^v\right) \cdot [(\ ) \circ s] = [(\ ) \not\vdash r] \cdot [(\ ) \circ s] = \ominus_r^s$. □.

This is an abstraction of the AND-process decomposition of clausal logic programs.

REFERENCES

[Benabou] BENABOU, J., Les distributeurs, Report no. 33, January 1973, Institute of Pure and Applied Mathematics, Catholic University of Louvain.

[Bernow] BERNOW, S. and RASKIN, P., Ecology of scientific consciousness, *Telos* **28**, Summer, 1976.

[Birkhoff] BIRKHOFF, G., *Lattice theory*, 3rd ed., American Mathematical Society Colloquium Publications, Vol. 25, American Mathematical Society, Providence, R. I., 1967, 325, 344. *MR* **37** #2638

[Girard] GIRARD, J. Y., Linear logic, *Theoret. Comp. Sci.* **50** (1987), 1–102; Technical Report, 1986, Equipe de Logique Mathématique, UER de Mathématiques, Université Paris VII.

[Henkin] HENKIN, L., MONK, J. D. and TARSKI, A., *Cylindric Algebras*, Part II, Studies in Logic and the Foundations of Mathematics, Vol. 115, North-Holland, Amsterdam–New York, 1985. *MR* **86m**: 03095b

[Hoare78] HOARE, C. A. R., Communicating sequential processes, *Comm. ACM* **21** (1978), 666–677.

[Hoare87] HOARE, C. A. R., HAYES, I. J., HE, JIFENG, MORGAN, C. C., ROSCOE, A. W., SANDERS, J. W., SORENSEN, I. H., SPIVEY, J. M. and SUFRIN, B. A., Laws of programming, *Comm. ACM* **30** (1987), 672–686. (Not in *MR*.)

[Hussey] HUSSEY, E., *The PreSocratics*, Scribner, 1972.

[Hyland] HYLAND, J. M. E., JOHNSTONE, P. T. and PITTS, A. M., Tripos theory, *Math. Proc. Cambridge Philos. Soc.* **88** (1980), 205–231. *MR* **81i:** 03102

[Kent87] KENT, R. E., Introduction to dialectical nets, 25th Allerton Conference on Communication, Control and Computing, Monticello, Illinois, 1987.

[Kent88] KENT, R. E., The logic of dialectical processes, 4th Workshop on Mathematical Foundations of Programming Semantics (Boulder, Colorado, 1988); Technical Report, 1989, Digital Systems Laboratory, Helsinki University of Technology, Espoo, Finland.

[Kent89] KENT, R. E., The standard aspect of dialectical logic (manuscript, submitted for publication).

[Lambek] LAMBEK, J., The mathematics of sentence structure, *Amer. Math. Monthly* **65** (1958), 154–170. *MR* **21** #4904

[Lawvere] LAWVERE, F. W., Adjointness in foundations, *Dialectica* **23** (1969), 281–296.

[Manes] MANES, E., Assertional categories, *Mathematical Foundations of Programming Language Semantics* (New Orleans, Louisiana, 1987), Lecture Notes in Comput. Sci., Vol. 298, Springer-Verlag, Berlin–New York, 1988, 85–120. *MR* **90f:** 68110

[Milner] MILNER, R., Calculi for synchrony and asynchrony, *Theoret. Comput. Sci.* **25** (1983), 267–310. *MR* **85g:** 68020

[Piccone] PICCONE, P., Dialectical logic today, *Telos* **1** (1968).

(Received August 10, 1988)

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
UNIVERSITY OF ILLINOIS AT CHICAGO
CHICAGO, IL 60680
U.S.A.

Current address:

DEPARTMENT OF COMPUTER AND
INFORMATION SCIENCE
UNIVERSITY OF ARKANSAS AT LITTLE ROCK
LITTLE ROCK, AR 72204
U.S.A.

Email: reKent@ualr.edu

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 63–70

К ТЕОРИИ ЭКСТРЕМАЛЬНЫХ ПОЛИНОМИАЛЬНЫХ ОПЕРАТОРОВ

Д. Л. БЕРМАН

1°. Пусть задан полином

$$t(x) = \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx). \tag{1}$$

Известно, что полином

$$\bar{t}(x) = \sum_{k=1}^{n} (a_k \sin kx - b_k \cos kx) \tag{2}$$

называется сопряженным по отношению к полиному (1). Множество всех полиномов вида (1) обозначим через Π_n. Обозначим через L множество всех суммируемых 2π-периодических функций. Введем линейное нормированное функциональное пространтсво E, обладающее свойствами: 1) элементы E суть функции из L; 2) если $f \in E$, то смещенная функция $f_t(x) = f(x+t)$ при любом $-\infty < t < \infty$ также из E, причем $\|f_t\| \leqq \|f\|$; 3) E содержит множество всех тригонометрических полиномов.

Важнейшим частным случаем пространства E является пространство C 2π-периодических непрерывных функций $f(x)$ с нормой $\|f(x)\| = \max |f(x)|$. Очевидно, что пространство L_p функций $f(x)$ периода 2π интегрируемых с p-ой степенью, $p \geqq 1$, с нормой

$$\|f(x)\|_{L_p} = \left(\int_0^{2\pi} |f(x)|^p dx \right)^{1/p}$$

также является пространством типа E.

Настоящая заметка примыкает к заметке [1], но в ней рассматривается сопряженная задача. Обозначим через $\tilde{\Omega}^{(r)}_{n,n+m}(E)$, где n, m, r

1980 *Mathematics Subject Classification* (1985 Revision). Primary 41A05; Secondary 41A35.

Key words and phrases. The set of trigonometric polynomials, extremal polynomial operator.

Akadémiai Kiadó, Budapest

— натуральные числа, множество всевозможных линейных операторов $U_{n,n+m}(f,x)$ из E в E, обладающих свойствами: 1) для любой $f\in E$, $U_{n,n+m}(f,x)\in\Pi_{n+m}$; 2) если $T\in\Pi_n$, то $U_{n,n+m}(T,x)=\tilde{T}^{(r)}(x)$, где $f^{(r)}(x)$ — производная порядка r от $f(x)$. Простейшей операцией из $\tilde{\Omega}^{(r)}_{n,n+m}(E)$ является выражение

$$\tau_n(f,x)=\frac{1}{m+1}\sum_{k=n}^{n+m}\tilde{S}_k^{(r)}(f,x),$$

где $S_k(f,x)$ — частная сумма порядка k ряда Фурье функции $f(x)$. Положим $\tilde{\varrho}^{(r)}_{n,n+m}=\inf\limits_{U_{n,n+m}\in\tilde{\Omega}^{(r)}_{n,n+m}(E)}\|U_{n,n+m}\|$. Пусть оператор $\tilde{U}\in$ $\in\tilde{\Omega}^{(r)}_{n,n+m}(E)$. Будем говорить, что он экстремальный в классе $\tilde{\Omega}^{(r)}_{n,n+m}$, если $\|\tilde{U}\|=\tilde{\varrho}^{(r)}_{n,n+m}$. Возникает естественный вопрос о нахождении в множестве операторов $\tilde{\Omega}^{(r)}_{n,n+m}(E)$ оператора с наименьшей нормой и о вычислении $\tilde{\varrho}^{(r)}_{n,n+m}$. Задача подобного рода была поставлена в [3]. Решить поставленную задачу при произвольном натуральном m видимо очень трудно. В настоящей заметке дается полное решение этой задачи для любых натуральных n и r и $m=n-1$.

2°. Теорема 1. *Для любого натурального r и любого $T\in\Pi_n$ имеет место тождество*

$$\tilde{T}^{(r)}(x)=\frac{1}{\pi}\int_{-\pi}^{\pi}T(x-t)\sin\Big(nt+\frac{r\pi}{2}\Big)\Big[n^r+2\sum_{k=1}^{n-1}k^r\cos(n-k)t\Big]dt. \tag{3}$$

Доказательство. Из (2) следует, что

$$\tilde{T}(x)=\frac{1}{\pi}\int_{-\pi}^{\pi}T(t)\sum_{k=1}^{n}\sin k(x-t)dt.$$

Поэтому

$$\tilde{T}^{(r)}(x)=\frac{1}{\pi}\int_{-\pi}^{\pi}T(t)\sum_{k=1}^{n}k^r\sin\Big[k(x-t)+\frac{r\pi}{2}\Big]dt.$$

Стало быть,

$$\tilde{T}^{(r)}(x)=\frac{1}{\pi}\int_{-\pi}^{\pi}T(x-t)\sum_{k=1}^{n}k^r\sin\Big(kt+\frac{r\pi}{2}\Big)dt. \tag{4}$$

С другой стороны, для любого $T\in\Pi_n$ выполняется равенство

$$(5)\qquad \frac{1}{\pi}\int_{-\pi}^{\pi} T(x-t)\sum_{k=1}^{n-1} k^r \sin\left[(2n-k)t+\frac{r\pi}{2}\right]dt=0.$$

Из (4) и (5) следует, что

$$(6)\qquad \begin{aligned}\tilde{T}^{(r)}(x)=\frac{1}{\pi}\int_{-\pi}^{\pi} T(x-t)\Big[n^r\sin\Big(nt+\frac{r\pi}{2}\Big)+\\ +\sum_{k=1}^{n-1}k^r\Big[\sin\Big(kt+\frac{r\pi}{2}\Big)+\sin\Big((2n-k)t+\frac{r\pi}{2}\Big)\Big].\end{aligned}$$

Так как $\sin\left(kt+\frac{r\pi}{2}\right)+\sin\left((2n-k)t+\frac{r\pi}{2}\right)=2\sin\left(nt+\frac{r\pi}{2}\right)\cos(n-k)t$, то из (6) вытекает (3).

3°. Построим теперь экстремальный оператор.

Теорема 2. *Оператор*

$$(7)\qquad \tilde{U}(f,x)=\frac{1}{\pi}\int_{-\pi}^{\pi} f(x-t)\sin\Big(nt+\frac{r\pi}{2}\Big)\Big[n^r+2\sum_{k=1}^{n-1}k^r\cos(n-k)t\Big]dt$$

принадлежит классу $\tilde{\Omega}^{(r)}_{n,2n-1}$.

Доказательство. Очевидно, что

$$\tilde{U}(f,x)=\frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin\Big[n(x-t)+\frac{r\pi}{2}\Big]\Big[n^r+2\sum_{k=1}^{n-1}k^r\cos(n-k)(x-t)\Big]dt.$$

Поэтому ясно, что оператор (7) переводит функции в тригонометрические полиномы порядка $2n-1$. Далее, в силу теоремы 1 для $T\in\Pi_n$, $\tilde{U}(T,x)=\tilde{T}^{(r)}(x)$. Итак, $\tilde{U}\in\tilde{\Omega}^{(r)}_{n,2n-1}$.

4°. Для дальнейшего нужна

Теорема 3. *Для всех* $t\in(-\infty,\infty)$ *выполняется неравенство*

$$(8)\qquad F_n(t)=n^r+2\sum_{k=1}^{n-1}k^r\cos(n-k)t\geqq 0.$$

Доказательство. Теорема 3 доказана в [1] и [6]. Поэтому мы здесь только наметим доказательство. Л. Фейер [2] доказал теорему: Пусть полином

$$T_n(x)=a_0+2a_1\cos t+\ldots+2a_n\cos nt$$

удовлетворяет условиям

$$a_\nu - 2a_{\nu+1} + a_{\nu+2} \geqq 0, \quad \nu = 0, 1, \ldots, (n-2), \; a_{n-1} - 2a_n \geqq 0, \; a_n \geqq 0.$$

Тогда $T_n(x) \geqq 0$, $-\infty < x < \infty$.

С помощью неравенств

$$\left(\frac{a+b}{2}\right)^k \leqq \frac{a^k + b^k}{2}, \quad k \geqq 1,$$

легко проверяется, что полином (8) удовлетворяет всем условиям теоремы Фейера. Поэтому выполняется (8).

Следствие 1. *При всех $t \in (-\infty, \infty)$ выполняется равенство*

$$\operatorname{sign}\left[\sin\left(nt + \frac{r\pi}{2}\right) F_n(t)\right] = \operatorname{sign} \sin\left(nt + \frac{r\pi}{2}\right), \tag{9}$$

за исключением корней полинома $F_n(t)$, где левая часть равенства (9) *равна нулю.*

5°. Для дальнейшего нужен аналог теоремы из [3].

Теорема 4. *Справедливо равенство*

$$\tilde{\varrho}^{(r)}_{n,2n-1}(C) = \inf_{\alpha_k, \beta_k} \mathcal{I}(\alpha_1, \ldots, \alpha_{n-1}, \beta_1, \ldots, \beta_{n-1}),$$

где

$$\mathcal{I} = \frac{1}{\pi} \int_{-\pi}^{\pi} \left| \tilde{D}^{(r)}_n(t) + \sum_{j=1}^{n-1} (\alpha_j \cos(n+j)t + \beta_j \sin(n+j)t) \right| dt, \quad \tilde{D}_n(t) = \sum_{\nu=1}^{n} \sin \nu t.$$

Если интеграл $\mathcal{I}$ достигает наименьшего значения при $\alpha_j = \alpha_j^{(1)}$, $\beta_j = \beta_j^{(1)}$, $j = 1, 2, \ldots, (n-1)$, то оператор

$$A(f, x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x-t) \left[\tilde{D}^{(r)}_n(t) + \sum_{j=1}^{n-1} (\alpha_j^{(1)} \cos(n+j)t + \beta_j^{(1)} \sin(n+j)t) \right] \tag{10}$$

является экстремальным в классе $\tilde{\Omega}^{(r)}_{n,2n-1}(C)$.

Так как $\tilde{D}^{(r)}_n$ нечетная функция при r четном и четная функция при r нечетном, то справедливы равенства

$$\begin{aligned} &\inf_{\alpha_k, \beta_k} \int_{-\pi}^{\pi} \left| \tilde{D}^{(2s)}_n(t) + \sum_{j=1}^{n-1} (\alpha_j \cos(n+j)t + \beta_j \sin(n+j)t) \right| dt = \\ &= \inf_{\gamma_k} \int_{-\pi}^{\pi} \left| \tilde{D}^{(2s)}_n(t) + \sum_{j=1}^{n-1} \gamma_j \sin(n+j)t \right| dt, \end{aligned} \tag{11}$$

$$\inf_{\alpha_k,\beta_k}\int_{-\pi}^{\pi}\left|\tilde{D}_n^{(2s+1)}(t)+\sum_{j=1}^{n-1}(\alpha_j\cos(n+j)t+\beta_j\sin(n+j)t\right|dt=$$

(12)

$$=\inf_{\delta_k}\int_{-\pi}^{\pi}\left|\tilde{D}_n^{(2s+1)}(t)+\sum_{j=1}^{n-1}\delta_j\cos(n+j)t\right|dt.$$

Воспользуемся теперь известными фактами теории наилучших приближений в метрике L [4]. Тогда из теоремы 4 и равенств (11), (12) получим

Теорема 5. 1) *Для того чтобы оператор* (10) *обладал наименьшей нормой в классе операторов* $\Omega_{n,2n-1}^{(2s)}$ *необходимо и достаточно, чтобы числа* $\{\gamma_i\}_{i=1}^{n-1}$ *из формулы* (11) *удовлетворяли условиям*

$$\int_0^{\pi}\operatorname{sign}\left[\tilde{D}_n^{(2s)}(t)+\sum_{j=1}^{n-1}\gamma_j\sin(n+j)t\right]\sin(n+i)t\,dt=0,\ i=1,2,\ldots,(n-1).$$

2) *Для того чтобы оператор* (10) *обладал наименьшей нормой в классе операторов* $\tilde{\Omega}_{n,2n-1}^{(2s+1)}$ *необходимо и достаточно, чтобы числа* $\{\delta_i\}_{i=1}^{n-1}$ *из формулы* (12) *удовлетворяли условиям*

$$\int_0^{\pi}\operatorname{sign}\left[\tilde{D}_n^{(2s+1)}(t)+\sum_{j=1}^{n-1}\delta_j\cos(n+j)t\right]\cos(n+i)t\,dt=0,\ i=1,2,\ldots,(n-1).$$

Поэтому из следствия 1 вытекает

Следствие 2. 1) *Для того чтобы оператор* (7) *обладал наименьшей нормой в классе операторов* $\tilde{\Omega}_{n,2n-1}^{(2s)}$ *необходимо и достаточно, чтобы выполнялись равенства*

$$\int_0^{\pi}\operatorname{sign}\sin nt\sin jt\,dt=0,\ j=n+1,\ldots,(2n-1). \tag{13}$$

2) *Для того чтобы оператор* (7) *обладал наименьшей нормой в классе операторов* $\tilde{\Omega}_{n,2n-1}^{(2s+1)}$ *необходимо и достаточно, чтобы выполнялись равенства*

$$\int_0^{\pi}\operatorname{sign}\cos nt\cos jt\,dt=0,\quad j=n+1,\ldots,(2n-1). \tag{14}$$

Теорема 6. *Среди всех линейных операторов $U_{n,2n-1}(f,x)$ из C в C, переводящих функции из C в полиномы порядка $2n-1$ и обладающих тем свойством, что для любого полинома T порядка не выше n имеет место равенство $U_{n,2n-1}(T,x)=T^{(r)}(x)$, оператор (7) обладает наименьшей нормой. При этом*

$$\tilde{\varrho}^{(r)}_{n,2n-1}=\frac{4}{\pi}n^r,\quad r=1,2,\dots. \tag{15}$$

Доказательство. В [4], стр. 99–100, доказано следующее утверждение. Пусть интегрируемая функция $\Phi(x)$ удовлетворяет условию $\Phi(x+\pi)=-\Phi(x)$. Пусть m и n — целые числа и отношение $\frac{m}{n}$ не есть нечетное число, тогда

$$\int_{-\pi}^{\pi} e^{imx}\Phi(nx)dx=0. \tag{16}$$

В частности, беря $\Phi(x)=\operatorname{sign}\sin x$ и $\Phi(x)=\operatorname{sign}\cos x$, получим из (16) равенства (13) и (14). Для вычисления $\tilde{\varrho}^{(r)}_{n,2n-1}$ заметим, что из формулы (7) следует, что

$$\tilde{\varrho}^{(r)}_{n,2n-1}=\frac{1}{\pi}\int_{-\pi}^{\pi}\left|\sin\left(nt+\frac{r\pi}{2}\right)\right|F_n(t)dt. \tag{17}$$

Так как

$$|\sin x|=\frac{2}{\pi}-\frac{4}{\pi}\sum_{k=1}^{\infty}\frac{\cos 2kx}{4k^2-1},$$

то

$$\left|\sin\left(nt+\frac{r\pi}{2}\right)\right|=\frac{2}{\pi}-\frac{4}{\pi}\sum_{k=1}^{\infty}(-1)^{rk}\frac{\cos 2knt}{4k^2-1}. \tag{18}$$

Из (17) и (18) выводим (15).

6°. Наряду с множеством операторов $\tilde{\Omega}^{(r)}_{n,2n-1}$ введем множество операторов $\Omega^{(r)}_{n,2n-1}$, которое определяется следующим образом. 1) Для любой $f\in C$ $U_{n,2n-1}(f,x)$ есть тригонометрический полином порядка не выше $2n-1$; 2) если $T\in\Pi_n$, то $U_{n,2n-1}(T,x)=T^{(r)}(x)$. Положим $\varrho^{(r)}_{n,2n-1}=\inf\limits_{U_{n,2n-1}\in\Omega^{(r)}_{n,2n-1}}\|U_{n,2n-1}\|$. В [1] доказано, что

$$\varrho^{(r)}_{n,2n-1}=\frac{4}{\pi}n^r,\quad r=1,2,\dots. \tag{19}$$

Из (15) и (19) вытекает, что

$$\tilde{\varrho}^{(r)}_{n,2n-1} = \varrho^{(r)}_{n,2n-1} = \frac{4}{\pi} n^r, \quad r = 1, 2, \ldots .$$

7°. Обозначим через $\mathcal{E}$ множество всех функциональных пространств типа E.

Теорема 7. *Имеет место равенство*

$$\sup_{E \in \mathcal{E}} \tilde{\varrho}^{(r)}_{n,2n-1}(E) = \tilde{\varrho}^{(r)}_{n,2n-1}(C) = \frac{4}{\pi} n^r.$$

Доказательство. Будем рассматривать оператор (7) как оператор из E в E. Обозначим его через $\tilde{U}(f, x, E)$. Было доказано, что он принадлежит классу $\tilde{\Omega}^{(r)}_{n,2n-1}(E)$. Известно [5] что для интеграла и нормы имеет место неравенство $\left\| \int f d\mu \right\| \leqq \int \|f\| d\mu$. Поэтому из (7) получим

$$\|\tilde{U}(f, x, E)\| \leqq \frac{1}{\pi} \int_{-\pi}^{\pi} \|f(x-t)\| \left| \sin\left(nt + \frac{r\pi}{2}\right) \right| F_n(t) dt. \tag{20}$$

Согласно определению пространство типа E, $\|f(x-t)\| \leqq \|f\|$. Стало быть, из (20) выводим, что

$$\|\tilde{U}(f, x, E)\| \leqq \|f\| \frac{1}{\pi} \int_{-\pi}^{\pi} \left| \sin\left(nt + \frac{r\pi}{2}\right) \right| F_n(t) dt. \tag{21}$$

Из доказательства теоремы 6 видно, что интеграл из правой части (21) равен $4n^r$. Поэтому в силу (21) получим, что для любого пространства типа E, $\|\tilde{U}\| \leqq \frac{4}{\pi} n^r$. Отсюда и из очевидного неравенства $\tilde{\varrho}^{(r)}_{n,2n-1}(E) \leqq \|\tilde{U}\|$ получим, что $\tilde{\varrho}_{n,2n-1}(E) \leqq \frac{4}{\pi} n^r$. Следовательно,

$$\sup_{E \in \mathcal{E}} \tilde{\varrho}_{n,2n-1}(E) \leqq \frac{4}{\pi} n^r. \tag{22}$$

С другой стороны, поскольку C пространство типа E, то

$$\sup_{E \in \mathcal{E}} \tilde{\varrho}^{(r)}_{n,2n-1}(E) \geqq \tilde{\varrho}^{(r)}_{n,2n-1}(C).$$

Согласно теореме 6 $\tilde{\varrho}^{(r)}_{n,2n-1}(C) = \frac{4n^r}{\pi}$. Стало быть,

$$\sup_{E \in \mathcal{E}} \tilde{\varrho}^{(r)}_{n,2n-1} \geqq \frac{4}{\pi} n^r. \tag{23}$$

Из (22) и (23) выводим теорему 7.

Замечание. Из теоремы 7, в частности, следует что пространство C является экстремальным среди пространств типа E, ибо неравенство

$$\tilde{\varrho}^{(r)}_{n,2n-1}(E) \leqq \frac{4}{\pi} n^r$$

при $E = C$ переходит в равенство.

Так как пространство L_p, $p \geqq 1$, является пространством типа E, то из теоремы 7 вытекает

$$\tilde{\varrho}^{(r)}_{n,2n-1}(L_p) \leqq \frac{4}{\pi} n^r, \quad p \geqq 1.$$

ЛИТЕРАТУРА

[1] Берман, Д. Л., Об одном экстремальном полиномиальном операторе дифференцирования, *Studia Sci. Math. Hungar.* **26** (1991), 105–110.

[2] Fejér, L., Über die Positivität von Summen, die nach trigonometrischen oder Legendreschen Funktionen fortschreiten I, *Acta Szeged* **2** (1925), 75–86. *Jahrbuch. Fortschritte Math.* **51**, 219

[3] Берман, Д. Л., Экстремальные задачи теории полиномиальных операторов, *Mat. Sb. (N.S.)* **60** (**102**) (1963), 354–365. *MR* **27** #516

[4] Ахиезер, Н. И., *Лекции по теории аппроксимации*, Наука, Москва, 1965. *MR* **32** #6108

[5] Данфорд, Н. и Шварц, Дж. Т., *Линейные операторы*, Издат. Иностр. Лит., Москва, 1962. *MR* **35** #7138

[6] Берман, Д. Л., Обобщение тождества Ф. Рисса и роль обобщенного тождества в теории линейных полиномиальных операций, *Kazan. Gos. Univ. Učen. Zap.* **127** (1967), 17–34. *MR* **43** #7845

(Поступило 10-ого января 1989. г.)

УЛ. ТУРКУ 18, КВ. 90
192238 САНКТ ПЕТЕРБУРГ
RUSSIA

Current address:

1357 W. ESTES # Q2
CHICAGO, IL 60626
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 71–77

INTEGRATION OF VECTOR-VALUED CONTINUOUS FUNCTIONS AND THE RIESZ REPRESENTATION THEOREM

L. A. KHAN

Abstract

We define and establish existence of the integral of functions in $C_b(X, E)$ with respect to certain E'-valued measures. We also obtain a Riesz representation type theorem.

1. Introduction

Let X be a completely regular Hausdorff space, E a real Hausdorff topological vector space (TVS) with non-trivial dual E', and $C_b(X, E)$ the vector space of all bounded continuous E-valued functions on X endowed with the uniform topology. In Section 2, we introduce some terminology and prove a density theorem. In Section 3, we define and establish existence of the integral of functions in $C_b(X, E)$ with respect to certain E'-valued measures defined on the algebra generated by zero sets of X. Finally, in Section 4, we obtain a Riesz representation theorem which characterizes the dual of a subspace of $C_b(X, E)$.

2. Terminology and preliminaries

Let $C_{tb}(X, E)$ (resp. $C_{rc}(X, E)$) denote the subspace of $C_b(X, E)$ consisting of those functions f such that $f(X)$ is totally bounded (relatively compact). When $E = \mathbb{R}$, the real field, we shall write $C_b(X)$ for $C_b(X, E)$. Let $C_b(X) \otimes E$ denote the vector space spanned by the set of all functions of the form $g \otimes a$, where $g \in C_b(X)$, $a \in E$, and $(g \otimes a)(x) = g(x)a$ $(x \in X)$. The *uniform topology* u on $C_b(X, E)$ is defined as the linear topology which has a base at 0 consisting of all sets of the form $\{f \in C_b(X, E)\colon f(X) \subseteqq W\}$, where W varies over a base $\mathcal{W}$ of neighbourhoods of 0 in E.

Following [6], [8], E is said to be *admissible* (resp. have the *approximation property*) if the identity map on E can be approximated uniformly on

1980 *Mathematics Subject Classification.* Primary 46E40, 46G10; Secondary 28A25.

Key words and phrases. Vector-valued bounded continuous functions, uniform topology, density theorem, vector-valued measures, existence of integral, Riesz representation theorem.

compact (totally bounded) sets by continuous (and linear) maps with range contained in finite dimensional subspaces of E.

We now prove the following density theorem which will justify somewhat our assumption in Section 4 that $C_b(X) \otimes E$ is u-dense in $C_{tb}(X, E)$.

THEOREM 2.1. (a) *E is admissible iff, for all topological spaces X, $C_b(X) \otimes E$ is u-dense in $C_{rc}(X, E)$.*

(b) *If E has the approximation property, then $C_b(X) \otimes E$ is u-dense in $C_{tb}(X, E)$.*

(c) *If X is a normal space of finite covering dimension* [3] *and E any TVS, then $C_b(X) \otimes E$ is u-dense in $C_{rc}(X, E)$.*

PROOF. (a) Suppose E is admissible, and let $f \in C_{rc}(X, E)$ and $W \in \mathcal{W}$ be balanced. Then there exists a continuous map $u: \overline{f(X)} \to E$ with range contained in a finite dimensional subspace of E such that $u(a) - a \in W$ for all $a \in \overline{f(X)}$. Then $h = u \circ f \in C_b(X) \otimes E$ and $h(x) - f(x) \in W$ for all $x \in X$.

Conversely, let $A \subseteq E$ be a compact set and $W \in \mathcal{W}$. Since, by hypothesis, $C_b(A) \otimes E$ is u-dense in $C_b(A, E)$, there exists some $v = \sum_{i=1}^{n} v_i \otimes a_i$ ($v_i \in C_b(A)$, $a_i \in E$) in $C_b(A) \otimes E$ such that $v(a) - a \in W$ for all $a \in A$. Note that the range of v is contained in the subspace spanned by $\{a_1, \ldots, a_n\}$. Thus E is admissible.

(b) Its proof is similar to the first part of (a).

(c) Since X is normal, its Stone–Čech compactification βX also has finite covering dimension ([3], p. 245). So, by [8], Theorem 1, $C_b(\beta X) \otimes E$ is u-dense in $C_b(\beta X, E)$. Note that each function in $C_b(X)$ or $C_{rc}(X, E)$ has a continuous extension to all of βX. Hence $C_b(X) \otimes E$ and $C_{rc}(X, E)$ are linearly isomorphic to $C_b(\beta X) \otimes E$ and $C_b(\beta X, E)$, respectively, and so the result follows.

We mention here that if E is a complete TVS, then $C_{rc}(X, E) = C_{tb}(X, E)$.

We next introduce some measure theoretic terminology. Let $B(X)$ denote the algebra generated by zero sets of X and $M(X)$ the space of all bounded real-valued finitely-additive regular (with respect to the family of zero sets) measures on $B(X)$. By a theorem of Aleksandrov (see [9], part I, Theorem 6), $(C_b(X), u)' = M(X)$ via the linear isometry $L \to \mu$, where $L(g) = \int_X g d\mu$ for all $g \in C_b(X)$. Following [9], a measure $\mu \in M(X)$ is called *σ-additive* if, for any sequence $\{Z_n\}$ of zero sets of X with $Z_n \downarrow \phi$, $|\mu|(Z_n) \to 0$; μ is called *τ-additive* if, for any net $\{Z_\lambda\}$ of zero sets of X with $Z_\lambda \downarrow \phi$, $|\mu|(Z_\lambda) \to 0$. If every σ-additive measure in $M(X)$ is τ-additive, then X is called *measure compact* [7].

For each $W \in \mathcal{W}$, let $M_W(X, E')$ denote the set of all finitely-additive E'-valued set functions m on $B(X)$ such that

(i) for each $a \neq 0$ in E, $m_a(F) = m(F)(a)$ ($F \in B(X)$) determines an element m_a of $M(X)$;

(ii) there exists a constant $r > 0$ such that $|m|_W(X) \leqq r$, where, for each $F \in B(X)$, we define $|m|_W$ by

$$|m|_W(F) = \sup \sum_i |m(F_i)(a_i)|,$$

(it would be the same for a balanced W) the supremum being taken over all finite partitions $\{F_i\}$ of F into sets in $B(X)$ (henceforth referred as a $B(X)$-partition) and all finite collections $\{a_i\} \subseteqq W$.

Let $M(X, E') = \bigcup_{W \in \mathcal{W}} M_W(X, E')$. By an argument similar to the one used in ([5], Lemma 2.3(i)), it is easy to verify that if $m \in M_W(X, E')$, then $|m|_W \in M(X)$.

3. The definition and existence of the integral

By Klee [6], we can always assume that $\mathcal{W}$ is a base consisting of closed balanced shrinkable neighbourhoods of 0 in E. (A neighbourhood W of 0 in a TVS is said to be *shrinkable* if $kW \subseteqq \operatorname{int} W$ for $0 \leqq k < 1$.) The advantage of taking such a base is that the Minkowski functional p_W of each $W \in \mathcal{W}$ is continuous and absolutely homogeneous, and further that $W = \{x \in X : p_W(x) \leqq 1\}$ ([6], Theorem 5). For this reason, our approach is simpler than [5] and [4] where the notions of F-seminorms and bipolars are used, respectively. If $f \in C_b(X, E)$ and $W \in \mathcal{W}$, we write $\|f\|_W = \|p_W \circ f\|$.

Let $m \in M_W(X, E')$ $(W \in \mathcal{W})$ and $f \in C_b(X, E)$. Let D be the collection of all $\alpha = \{F_1, \ldots, F_n; x_1, \ldots, x_n\}$, where $\{F_i\}$ $(1 \leqq i \leqq n)$ is a $B(X)$-partition of X and $x_i \in F_i$. If $\alpha_1, \alpha_2 \in D$, define $\alpha_1 \geqq \alpha_2$ iff each set which appears in α_1 is contained in some set in α_2. In this way, D becomes an indexing set. Let $S_\alpha = \sum_{i=1}^{n} m(F_i) f(x_i)$. We now define the *integral* of f with respect to m over X by

$$\int_X dmf = \lim_{\alpha \in D} S_\alpha.$$

Regarding the conditions under which this integral exists, we obtain the following lemma.

LEMMA 3.1. *The integral* $\int_X dmf$, *defined above, exists in each of the following cases:*

(a) $f \in C_{tb}(X, E)$;

(b) $f \in C_b(X, E)$ *and* $|m|_W$ *is* τ*-additive;*

(c) $f \in C_b(X, E)$, $|m|_W$ *is* σ*-additive, and the range of* f *is measure compact.*

PROOF. We need to show that, in each case, $\{S_\alpha : \alpha \in D\}$ is a Cauchy net in $\mathbf{R}$.

(a) Let $\varepsilon > 0$. Choose a balanced shrinkable $V \in \mathcal{W}$ such that $V + V \subseteqq \varepsilon W$. Since $f(X)$ is totally bounded, there exist $y_1, \ldots, y_n \in X$ such that $f(X) \subseteqq \bigcup_{i=1}^{n} (f(y_i) + V)$. Let $V_i = \{x \in X : p_V(f(x) - f(y_i)) \leqq 1\}$. Then each $V_i \in B(X)$. Let $G_1 = V_1$ and $G_i = V_i \setminus \bigcup_{j=1}^{i-1} V_j$ $(2 \leqq i \leqq n)$. By keeping those G_i's which are non-empty, we get, $\{G_1, \ldots, G_k\}$ say, a $B(X)$-partition of X. Choose $x_i \in G_i$ and let $\alpha_0 = \{G_1, \ldots, G_k; x_1, \ldots, x_k\}$. Note that if x, y are in the same G_i, then $f(x) - f(y) \in \varepsilon W$. Then, for $\alpha_1, \alpha_2 \geqq \alpha_0$, we have

$$|S_{\alpha_1} - S_{\alpha_2}| \leqq |S_{\alpha_1} - S_{\alpha_0}| + |S_{\alpha_0} - S_{\alpha_2}|.$$

Suppose $\alpha_1 = \{F_1, \ldots, F_q; z_1, \ldots, z_q\}$, where each F_j is contained in some G_i and $z_j \in F_j$. Now

$$|S_{\alpha_1} - S_{\alpha_0}| = \left|\sum_{j=1}^{q} m(F_j) f(z_j) - \sum_{i=1}^{k} m(G_i) f(x_i)\right| =$$

$$= \left|\sum_{j=1}^{q} m(F_j) f(z_j) - \sum_{i=1}^{k} \sum_{\substack{j \\ F_j \subseteqq G_i}} m(F_j) f(x_i)\right| =$$

$$= \varepsilon \left|\sum_{i=1}^{k} \sum_{\substack{j \\ F_j \subseteqq G_i}} m(F_j)(\varepsilon^{-1}(f(z_j) - f(x_i)))\right| \leqq \varepsilon |m|_W(X).$$

Similarly, we can prove that $|S_{\alpha_2} - S_{\alpha_0}| \leqq \varepsilon |m|_W(X)$. Thus $|S_{\alpha_1} - S_{\alpha_2}| \leqq 2\varepsilon |m|_W(X)$, and consequently $\{S_\alpha : \alpha \in D\}$ is a Cauchy net in $\mathbf{R}$.

(b) Let $\varepsilon > 0$, and choose an open balanced shrinkable $V \in \mathcal{W}$ with $V + V \subseteqq \varepsilon W$. The collection $\mathcal{V} = \{f^{-1}(f(y) + V) : y \in X\}$ is a cover of X consisting of cozero sets. Labelling $\mathcal{V}$ as $\{V_\lambda : \lambda \in I\}$, we make I into a directed set by saying that $\lambda \geqq \gamma$ iff $V_\lambda \subseteqq V_\gamma$. By the τ-additivity of $|m|_W$, there exist $y_1, \ldots, y_k \in X$ such that $|m|_W\left(X \setminus \bigcup_{i=1}^{k} V_{\lambda_i}\right) < \varepsilon$, where

$$V_{\lambda_i} = f^{-1}(f(y_i) + V) = \{x \in X : p_V(f(x) - f(y_i)) < 1\}.$$

Define $G_1 = V_{\lambda_1}$, $G_i = \left(V_{\lambda_i} \setminus \bigcup_{j=1}^{i-1} V_{\lambda_j}\right)$ $(2 \leqq i \leqq k)$, and $G_{k+1} = X \setminus \bigcup_{i=1}^{k} V_{\lambda_i}$. Assuming that G_i's are non-empty, choose $x_i \in G_i$ and let $\alpha_0 = \{G_1, \ldots, G_{k+1}; x_1, \ldots, X_{k+1}\}$.

Let $\alpha_1, \alpha_2 \geqq \alpha_0$. Suppose $\alpha_1 = \{F_1, \ldots, F_q; z_1, \ldots, z_q\}$, where each F_j is contained in some G_i and $z_j \in F_j$. Then

$$|S_{\alpha_1} - S_{\alpha_0}| \leqq \Big|\sum_{i=1}^{k} \sum_{\substack{j \\ F_j \subseteqq G_i}} m(F_j)(f(z_j) - f(x_i))\Big| + \Big| \sum_{\substack{j \\ F_j \subseteqq G_{k+1}}} m(F_j) f(z_j)\Big| +$$

$$+ \Big| \sum_{\substack{j \\ F_j \subseteqq G_{k+1}}} m(F_j) f(x_{k+1})\Big| \leqq \varepsilon(|m|_W(X) + 2\|f\|_W).$$

Similarly, we obtain $|S_{\alpha_2} - S_{\alpha_0}| \leqq \varepsilon(|m|_W(X) + 2\|f\|_W)$. Consequently, $\{S_\alpha : \alpha \in D\}$ is a Cauchy net in $\mathbf{R}$.

(c) Suppose $|m|_W$ is σ-additive, and let $\mu(A) = |m|_W(f^{-1}(A))$ for every Baire set A of $f(X)$. Then μ is also σ-additive. Now, taking $\mathcal{V}$ as in part (b) and using the fact that μ is τ-additive (since $f(X)$ is measure compact), we can complete the proof by the argument of part (b).

REMARK. Parts (b) and (c) were proved in ([1]; [2], Lemma 3.11) assuming E a normed space.

LEMMA 3.2. *Let $m \in M_W(X, E')$ $(W \in \mathcal{W})$ and $f \in C_b(X, E)$. If the integral $\int_X dmf$ exists, then*

$$\Big|\int_X dmf\Big| \leqq \Big|\int_X (p_W \circ f) d|m|_W\Big| \leqq \|f\|_W |m|_W(X).$$

PROOF. It is a slight modification of [5], Lemma 2.3(ii) and is therefore omitted.

It is easy to verify that, if $m \in M(X, E')$, $g \in C_b(X)$, and $a \in E$, then $\int_X dm(g \otimes a) = \int_X g\, dm_a$.

4. The Riesz representation theorem

In this section we characterize the u-dual of $C_{tb}(X, E)$ via the integral representation. Throughout we assume that $C_b(X) \otimes E$ is u-dense in $C_{tb}(X, E)$.

THEOREM 4.1. *$(C_{tb}(X, E), u)' = M(X, E')$ via the linear isomorphism $L \to m$ given by*

$$L(f) = \int_X dmf \qquad (f \in C_{tb}(X, E)). \tag{1}$$

Furthermore, if L is represented as in (1) *with* $m \in M_W(X, E')$ $(W \in \mathcal{W})$, *then* $\|L\|_W = |m|_W(X)$, *where*

$$\|L\|_W = \sup\{|L(f)|: f \in C_{tb}(X, E), \|f\|_W \leqq 1\}.$$

PROOF. Let $m \in M_W(X, E')$ $(W \in \mathcal{W})$, and suppose that L is the linear functional on $C_{tb}(X, E)$ defined by (1). Then, by Lemma 3.2,

$$|L(f)| \leqq \|f\|_W |m|_W(X) \leqq |m|_W(X)$$

whenever $f \in C_{tb}(X, E)$ with $f(X) \subseteq W$. Hence $L \in (C_{tb}(X, E), u)'$. We now show that $\|L\|_W = |m|_W(X)$. Clearly, $\|L\|_W \leqq |m|_W(X)$. For any $\varepsilon > 0$, there exist a $B(X)$-partition $\{F_1, \ldots, F_n\}$ of X and a collection $\{a_1, \ldots, a_n\} \subseteq W$ such that

$$|m|_W(X) \leqq \Big|\sum_{i=1}^{n} m(F_i)(a_i)\Big| + \varepsilon.$$

Using the regularity of each m_{a_i} and complete regularity of X, as in [10], Lemma 4, we easily obtain

$$\Big|\sum_{i=1}^{n} m_{a_i}(F_i)\Big| \leqq \|L\|_W + \varepsilon.$$

Thus $|m|_W(X) \leqq \|L\|_W$.

Conversely, suppose that $L \in (C_{tb}(X, E), u)'$. Then there exist a $W \in \mathcal{W}$ and $r > 0$ such that

$$|L(f)| \leqq r\|f\|_W \qquad (f \in C_{tb}(X, E)). \tag{2}$$

For each $a \neq 0$ in E, let $L_a(g) = L(g \otimes a)$ $(g \in C_b(X))$. By (2), we have $|L_a(g)| \leqq r\|g\|p_W(a)$ for all $g \in C_b(X)$, and so $L_a \in (C_b(X), u)'$. Hence, by Aleksandrov's theorem, there exists an $m_a \in M(X)$ such that $L_a(g) = \int_X g dm_a$ $(g \in C_b(X))$ and $\|L_a\| = |m_a|(X)$.

For each $F \in B(X)$, define $m(F)(a) = m_a(F)$ $(a \in E)$. Then $|m(F)(a)| \leqq |m_a|(X) \leqq rp_W(a)$, and consequently m is a finitely additive E'-valued set function on $B(X)$. Using again the argument of [10], Lemma 4 and also (2), it follows that $|m|_W(X) \leqq r$. Thus $m \in M_W(X, E')$.

Now, for any $f = \sum_{i=1}^{k} f_i \otimes a_i$ $(f_i \in C_b(X), a_i \in E)$ in $C_b(X) \otimes E$,

$$L(f) = \sum_{i=1}^{k} L(f_i \otimes a_i) = \sum_{i=1}^{k} \int_X f_i dm_{a_i} = \sum_{i=1}^{k} \int_X dm(f_i \otimes a_i) = \int_X dmf.$$

Since, by our assumption, $C_b(X) \otimes E$ is u-dense in $C_{tb}(X, E)$, the above holds for all $f \in C_{tb}(X, E)$. This completes the proof.

REFERENCES

[1] FONTENOT, R. A., Strict topologies for vector-valued functions, *Canad. J. Math.* **26** (1974), 841–853. *MR* **50** #961

[2] FONTENOT, R. A., "Strict topologies for vector-valued functions": Corrigendum, *Canad. J. Math.* **27** (1975), 1183–84. *MR* **52** #8894

[3] GILLMAN, L. and JERISON, M., *Rings of continuous functions*, The University Series in Higher Mathematics, D. Van Nostrand Co., Inc., Princeton, N. J., 1960. *MR* **22** #6994

[4] KATSARAS, A. K., On the strict topology in the nonlocally convex setting II, *Acta Math. Hungar.* **41** (1983), 77–88. *MR* **85f:** 46049

[5] KHAN, L. A. and ROWLANDS, K., On the representation of strictly continuous linear functionals, *Proc. Edinburgh Math. Soc.* (2) **24** (1981), 123–130. *MR* **83b:** 46050

[6] KLEE, V., Shrinkable neighborhoods in Hausdorff linear spaces, *Math. Ann.* **141** (1960), 281–285. *MR* **24** #A1003

[7] MORAN, W., Measures and mappings on topological spaces, *Proc. London Math. Soc.* (3) **19** (1969), 493–508. See *Zbl* **183**, 372

[8] SHUCHAT, A. H., Approximation of vector-valued continuous functions, *Proc. Amer. Math. Soc.*, **31** (1972), 97–103. *MR* **44** #7267

[9] VARADARAJAN, V. S., Measures on topological spaces, *Mat. Sb.* (N.S.) **55** (**97**) (1961), 35–100. *MR* **26** #6342. See also *Amer. Math. Soc. Transl.* (2) **48** (1965), 161–228.

[10] WELLS, J., Bounded continuous vector-valued functions on a locally compact space, *Michigan Math. J.* **12** (1965), 119–126. *MR* **31** #593

(Received April 23, 1989)

DEPARTMENT OF MATHEMATICS
QUAID-I-AZAM UNIVERSITY
PAK–45320 ISLAMABAD
PAKISTAN

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 79–86

A GENERALIZATION OF THE RICCATI EQUATION

I. BIHARI

1. Consider the ordinary second order linear differential equation

$$(1) \qquad (py')'+qy=0, \quad x\in J[x_0,\infty),\ p>0,\ p,q\in C(J), \quad \int_{x_0}^{\infty} p^{-1}=\infty.$$

The function $\xi=\frac{py'}{y}$ — where y is a solution of (1) — satisfies the Riccati equation

$$(2) \qquad \xi'+\frac{\xi^2}{p}+q=0$$

with the restriction $y\neq 0$, thus the estimates obtained by (2) concerning y will also be restricted, while in the present paper results without such restriction will be given, both for oscillatory and nonoscillatory cases, too.

This defect (disadvantage, insufficiency, inadequacy) of the Riccati equation can be eliminated or we can get rid of it as it will be done in the sequel. It is well known that the function

$$(3) \qquad A[y]=py^2+\frac{y'^2}{q}$$

of y — called amplitude of y — plays an important role in many investigations (oscillation, asymptotic behaviour, stability, some monotonicity problems, etc.). Another important function is

$$(4) \qquad B[y_1,y_2]=y_1^2+y_2^2=\eta(x)=\eta$$

— where (y_1,y_2) is a pair of linearly independent solutions of (1) — called the "weight" of the pair (y_1,y_2).

Let us deduce a Riccati-like equation concerning

$$z=\frac{p\eta'}{\eta}$$

1990 *Mathematics Subject Classification.* Primary 34A30; Secondary 34C11, 34E05.
Key words and phrases. Riccati equation, estimate of solutions, qualitative theory.

Akadémiai Kiadó, Budapest

which has the advantage — compared to (2) — that η never vanishes. We have in turn

$$z' = \frac{(p\eta')'\eta - p\eta'^2}{\eta^2} = \frac{(p\eta')'}{\eta} - \frac{1}{p}\left(\frac{p\eta'}{\eta}\right)^2$$

or

$$z' + \frac{z^2}{p} = \frac{(p\eta')'}{\eta}.$$

Here

$$\begin{gathered}(p\eta')' = 2[p(y_1y_1' + y_2y_2')]' = \\ = 2[(py_1')'y_1 + (py_2')'y_2 + p(y_1'^2 + y_2'^2)] = \\ = 2[-q(y_1^2 + y_2^2) + p(y_1'^2 + y_2'^2)]\end{gathered}$$

and

$$\frac{(p\eta')'}{\eta} = -2q + \frac{2p\Delta}{\eta}, \qquad \Delta = y_1'^2 + y_2'^2.$$

Thus

$$z' + \frac{z^2}{p} + 2q = \frac{2p\Delta}{\eta}.$$

This is not sufficient for our purposes, since we cannot do with the term $\frac{2p\Delta}{\eta}$. Therefore we proceed as follows. Applying the identity

$$(u_1v_1 + u_2v_2)^2 + (u_1v_2 - u_2v_1)^2 = (u_1^2 + u_2^2)(v_1^2 + v_2^2)$$

to

$$u_i = y_i, \quad v_i = y_i', \quad i = 1, 2$$

we get

$$(y_1y_1' + y_2y_2')^2 + (y_1y_2' - y_2y_1')^2 = (y_1^2 + y_2^2)(y_1'^2 + y_2'^2)$$

or

$$\frac{\eta'^2}{4} + W^2 = \eta\Delta, \qquad W = y_1'y_2 - y_2'y_1 = \frac{c}{p}, \qquad c = \text{const}$$

whence

$$\frac{2p\Delta}{\eta} = \frac{z^2}{2p} + \frac{2c^2}{p\eta^2},$$

finally

$$z' + \frac{z^2}{2p} + 2q = \frac{2c^2}{p\eta^2}, \tag{5}$$

which is the Riccati-type equation looked for and valid without any restriction on the values of $y(t)$. The last remark holds also concerning the following estimates involved by (5)

$$z < z_0 - 2\int_{x_0}^{x}(q(s) - c^2p^{-1}(s)\eta^{-2}(s))ds \stackrel{\text{def}}{=} f(x), \quad z_0 = z(x_0), \tag{6}$$

$$\eta < \eta_0 \exp\left(\int_{x_0}^{x} f(s)p^{-1}(s)ds\right) \stackrel{\text{def}}{=} F(x), \tag{7}$$

$$z' + \frac{f^2}{2p} + 2q > \frac{2c^2}{pF^2}, \tag{8}$$

$$z > z_0 - \int_{x_0}^{x}\left(\frac{f^2}{2p} + 2q - \frac{2c^2}{pF^2}\right)dx \stackrel{\text{def}}{=} g(x), \tag{9}$$

$$\eta > \eta_0 \exp\left(\int_{x_0}^{x} gp^{-1}dx\right) \stackrel{\text{def}}{=} G(x), \tag{10}$$

$$G(x) < \eta < F(x). \tag{11}$$

These estimates have another insufficiency. Namely the term

$$I(x) = \int_{x_0}^{x} p^{-1}(s)\eta^{-2}(s)ds$$

involved in the above formulas is unknown in general, since it involves η.

However, (6)–(11) can be useful provided the integrals

$$I(x), \quad I_1(x) = \int_{x_0}^{x} q(s)ds, \quad \text{or} \quad c^2I - I_1$$

are convergent or at least bounded as $x \to \infty$. I_1 must be bounded from below, too.

In general, nothing is known in this respect, but some important particular cases were studied earlier:

(A) If (1) is non-oscillatory, suppose

$$-\infty < \liminf_{x\to\infty} I_1 < \limsup_{x\to\infty} I_1 < \infty \tag{12}$$

which is satisfied if $q \geqq 0$. Namely, then $\lim_{x\to\infty} I_1$ exists [1]. Furthermore there is a "small" solution y_1 with

$$I_2 = \int^{\infty} p^{-1} y_1^{-2} = \infty$$

and for every other solution y_2 — which is linearly independent of y_1 — called "large" solution — we have

$$I_3 = \int^{\infty} p^{-1} y_2^{-2} < \infty$$

involving

$$I_4 = \int^{\infty} p^{-1}\eta^{-1} < \infty, \quad \eta = y_1^2 + y_2^2$$

(see [2, p. 355]). Now we show that $I(x)$ is also convergent as $x \to \infty$ provided $q > 0$ (or $q < 0$). Clearly, in this case y_1 and y_2 are convex (concave), non-vanishing and monotonic for $x > x_1$ with some $x_1 > 0$ and $\lim_{x\to\infty} y_2$ exists, $\lim_{x\to\infty} y_2^2$, too, and is finite or infinite, but by all means greater than ε, where $\varepsilon > 0$ is a fixed number. This follows from the convergence of I_3. On the other hand if y_2 is concave it has to be increasing for $x > x_1$, having no zero greater than x_1, i.e., again $\lim_{x\to\infty} y_2^2 > \varepsilon$, implying $\eta > \varepsilon$ and the convergence of $I(x)$ as $x \to \infty$. — The case $q < 0$ can be treated in the same way.

By the way, let it be remarked that if we drop the condition $q \geqq 0$ (or $q \leqq 0$) *throughout* J then we can construct such a q that I_4 is convergent and I is not (see Example 3).

Thus in the present case $z < k = \text{const.}$ and the estimates (6)–(11) make sense.

What is about the accuracy of these estimates? Of course, the omitted term $\frac{z^2}{2p}$ of (5) must be small compared to $c^2 p^{-1}\eta^{-2} - q$ as $x \to \infty$, i.e.

$$R = \frac{c^2 p^{-1}\eta^{-2} - q}{\frac{z^2}{p}} = \frac{c^2\eta^{-2} - pq}{z^2}$$

must be large compared to 1 as $x \to \infty$.

(B) If (1) is oscillatory then in general such estimates cannot be obtained. However, there are examples where both I_1 and I (or $c^2 I - I_1$) are convergent and the above estimates can be carried out. An example where I_1 and I are not convergent but $c^2 I - I_1$ converges is as follows.

EXAMPLE 1. As it is well-known the Bessel function $J_\nu(x)$ with the index $\nu = \frac{3}{2}$ — or more precisely — the functions

$$y_1 = \sqrt{x} J_\nu(x) = \sqrt{\frac{2}{\pi}}\left(\frac{\sin x}{x} - \cos x\right)$$
$$y_2 = \sqrt{x} J_{-\nu}(x) = \sqrt{\frac{2}{\pi}}\left(-\frac{\cos x}{x} - \sin x\right) \qquad \nu = \frac{3}{2}$$

satisfy the equation

$$y'' + \left(1 - \frac{2}{x^2}\right) y = 0.$$

Here

$$\eta = x(J_\nu^2 + J_{-\nu}^2) = \frac{2}{\pi}\left(\frac{1}{x^2} + 1\right), \quad z = \frac{\eta'}{\eta} = -\frac{2}{x(x^2+1)}, \quad c = \frac{2}{\pi}$$

$$\frac{c^2}{\eta^2} - q = \frac{3x^2+2}{x^2(x^2+1)^2}, \quad c^2 I - I_1 \text{ is convergent}$$

as $x \to \infty$ and

$$R = \frac{3x^2+2}{4} \to \infty \quad \text{as} \quad x \to \infty.$$

Therefore the approximations above are good enough.

EXAMPLE 2. If in equation

$$y'' + \frac{1}{\lambda x^2} y = 0, \quad \lambda > 0, \quad x_0 > 0$$

1° $\lambda > 4$, then $y_1 = \sqrt{x} x^\mu$, $y_2 = \sqrt{x} x^{-\mu}$, $\mu = \frac{1}{2}\sqrt{\frac{\lambda - 4}{\lambda}}$

$$\eta = x(x^{2\mu} + x^{-2\mu}), \quad c = 2\mu, \quad \lim_{x\to\infty} R = -\frac{\sqrt{\lambda} + \sqrt{\lambda - 4}}{16},$$

2° $\lambda = 4$, then $y_1 = \sqrt{x}$, $y_2 = \sqrt{x}\log x$, $\eta = x(1 + \log^2 x)$, $c = -1$

$$\lim_{x\to\infty} R = -\frac{1}{\lambda} = -\frac{1}{4}.$$

3° $0 < \lambda < 4$, then $y_1 = \sqrt{x}\cos(\nu \log x)$, $y_2 = \sqrt{x}\sin(\nu \log x)$

$$\nu = \frac{1}{2}\sqrt{\frac{4 - \lambda}{\lambda}}, \quad \eta = x, \quad c = -\nu, \quad \lim_{x\to\infty} R = -\frac{1}{4}.$$

In all the three cases I_1 and I are convergent, the estimates make sense, but are not good, since R is not large as $x \to \infty$.

We can raise the problem of finding the value of the parameter κ which furnishes the maximum of R as $x \to \infty$ when we replace — at fixed (y_1, y_2) — η by $\overline{\eta} = \kappa y_1^2 + y_2^2$. In Example 1 this value is $\kappa = 1$.

There is another way to establish the accuracy of (6)–(11). The two limits

$$\lim_{x\to\infty} \int_{x_0}^{x} \frac{z^2}{p}, \qquad \lim_{x\to\infty} \frac{1}{X} \int_{x_0}^{X} \left(\int_{x_0}^{x} \left(\frac{c^2}{p\eta^2} - q \right) ds \right) dx$$

exist or do not exist at the same time as it can be proved in the wake of a theorem by P. Hartman [2, p. 365]. This means that $\frac{z^2}{p}$ is "integral small" instead to be small in common sense.

2. In the non-oscillatory case, analysing the problem of convergence of the integral $I(x)$ as $x \to \infty$ we need some formulas expressing y_1, y_2 and q by η. For the sake of simplicity assume $p = 1$. If y_1 is a solution of (1) with $y_1(x_0) = 1$, $y'(x_0) = 0$, then

$$y_2 = y_1 \int_{x_0}^{x} \frac{ds}{y_1^2}$$

is a solution with $y_2(x_0) = 0$, $y_2'(x_0) = 1$ and y_1, y_2 are linearly independent and

$$\eta = y_1^2 + y_2^2 = y_1^2 \left[1 + \left(\int_{x_0}^{x} y_1^{-2}(s) ds \right)^2 \right]. \tag{13}$$

Let $\int_{x_0}^{x} y_1^{-2} ds = u(x)$, then $\eta = \frac{1}{u'}(1 + u^2)$, whence

$$\frac{u'}{1+u^2} = \frac{1}{\eta}, \quad \operatorname{arctg} u = \int_{x_0}^{x} \eta^{-1} dx + C.$$

Thus

$$u = \int_{x_0}^{x} y_1^{-2} ds = \operatorname{tg}\left(\int_{x_0}^{x} \frac{ds}{\eta(s)} + C \right). \tag{14}$$

Hence

$$\frac{1}{y_1^2} = \frac{1}{\eta \cos^2 \left(\int \frac{dx}{\eta} + C \right)}, \quad y_1 = \sqrt{\eta} \cos\left(\int_{x_0}^{x} \frac{ds}{\eta(s)} + C \right)$$

and

$$y_2 = \sqrt{\eta - y_1^2} = \sqrt{\eta}\sin\left(\int_{x_0}^{x} \frac{ds}{\eta(s)} + C\right).$$

But

$$\begin{aligned} y_1(x_0) &= \sqrt{\eta_0}\cos C = 1, \\ y_2(x_0) &= \sqrt{\eta_0}\sin C = 0, \end{aligned} \qquad \eta_0 = 1$$

giving $C = 0$. Since $\int_{x_0}^{\infty} y_1^{-2} = \infty$ (being y_1 the "small solution") (14) implies $\int_{x_0}^{\infty} \eta^{-1}dx = \frac{\pi}{2}$ involving $\int_{x_0}^{x} = \frac{\pi}{2} - \int_{x}^{\infty}$. Finally,

$$y_1 = \sqrt{\eta}\sin\left(\int_{x}^{\infty} \frac{ds}{\eta(s)}\right), \quad y_2 = \sqrt{\eta}\cos\left(\int_{x}^{\infty} \frac{ds}{\eta(s)}\right).$$

The coefficient q can be expressed from equation

$$z' + \frac{z^2}{2} + 2q = 2\eta^{-2}, \quad z = \frac{\eta'}{\eta}$$

obtaining

$$q = \frac{4 + \eta'^2 - 2\eta\eta''}{4\eta^2}.$$

Now we can construct an η (i.e. an equation (1)) with

$$\lim_{x\to\infty} I_2 < \infty, \qquad \lim_{x\to\infty} I = \infty.$$

A function like $\frac{1}{x^{1+\alpha}}$ $(\alpha > 0)$ or $\frac{1}{x\log^2 x}$ cannot be chosen as $\frac{1}{\eta(x)}$, since then both I_2 and I are convergent. Therefore such a function has to be modified as follows.

EXAMPLE 3. Let

$$S(x) = \sum_{n=1}^{\infty} \frac{1}{(2a)^n} x^n,$$

then

$$S(a) = \sum \frac{a^n}{(2a)^n} = \sum \frac{1}{2^n} < \infty,$$

but

$$S(a^2) = \sum \frac{a^{2n}}{(2a)^n} = \sum \left(\frac{a}{2}\right)^n = \infty,$$

provided $a \geqq 2$. Now let us construct a sequence of functions $f_n(x)$ the graph of which are triangles having as basis the intervals $\left(n, n+\frac{1}{(2a)^n}\right)$ of the x axis and the heights a^n $(n = 1, 2, \dots)$ and $f_n(x) = 0$ elsewhere, then by adding $f_n(x)$ to one of the above functions we obtain a function η^{-1} with

$$\int^{\infty} \frac{1}{\eta} < \infty, \qquad \int^{\infty} \frac{1}{\eta^2} = \infty.$$

REFERENCES

[1] WINTNER, A., A criterion of oscillatory stability, *Quart. Appl. Math.* **7** (1949), 115–117. *MR* **10**–456

[2] HARTMAN, PH., *Ordinary differential equations*, John Wiley and Sons, New York–London, 1964. *MR* **30** #1270

(Received May 8, 1989)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 87–97

HERMITE-FEJÉR INTERPOLATIONS OF HIGHER ORDER. III

R. SAKAI and P. VÉRTESI[1]

1. Introduction

Let $X = \{x_{kn} = \cos \vartheta_{kn}\} \subset [-1, 1]$,

$$-1 \leqq x_{nn} < x_{n-1,n} < \cdots < x_{2n} \leqq x_{1n} \leqq 1, \quad n = 1, 2, \dots , \tag{1.1}$$

be an infinite triangular interpolatory matrix. The unique Hermite–Fejér interpolatory polynomial $H_{nm}(f, X, x)$ of order $\leqq mn - 1$ ($m \geqq 1$, fixed integer) for an arbitrary continuous function $f(x)$ in $[-1, 1]$ ($f \in C$, shortly) is defined by

$$H_{nm}^{(t)}(f, X, x_{kn}) = \delta_{0t} f(x_{kn}), \quad k = 1, 2, \dots , n, \quad t = 0, 1, \dots , m - 1. \tag{1.2}$$

By (1.2) and some obvious short notations H_{nm} can be written as

$$H_{nm}(f, x) = \sum_{k=1}^{n} f(x_k) h_{knm}(x), \quad n = 1, 2, \dots , \tag{1.3}$$

where for the polynomials h_k of degree exactly $mn - 1$

$$h_k^{(t)}(x_j) = \delta_{0t} \delta_{kj}, \quad 1 \leqq k, j \leqq n, \ 0 \leqq t \leqq m - 1. \tag{1.4}$$

An explicit form of h_k is

$$h_{knm}(x) = l_{kn}^m(x) \sum_{i=0}^{m-1} e_{iknm} (x - x_k)^i, \quad 1 \leqq k \leqq n, \tag{1.5}$$

1980 *Mathematics Subject Classification* (1985 Revision). Primary 41A05.
Key words and phrases. Interpolation, Hermite–Fejér parabolas.

[1]This author's research was supported by the Hungarian National Foundation for Scientific Research Grants nos. 1910 and T7570.

Akadémiai Kiadó, Budapest

where l_k are the Lagrange fundamental polynomials of degree exactly $n-1$, i.e. with $\omega_n(x) := c_n \prod_{k=1}^{n} (x - x_k)$

$$l_{kn}(x) = \frac{\omega_n(x)}{\omega_n'(x_k)(x - x_k)}, \quad 1 \leqq k \leqq n,$$

the coefficients e_{ik} can be obtained by (1.4) (cf. R. Sakai [1] or P. Vértesi [2]).

When $X = X^{(\alpha,\beta)}$, i.e. when the nodes (1.1) are the roots of the n-th Jacobi polynomials $P_n^{(\alpha,\beta)}(x), \alpha, \beta > -1$ (cf. G. Szegő [3; 4.1]) as a detailed analysis shows the behaviour of the polynomials $H_{nm}^{(\alpha,\beta)}(f, x)$ resembles the Lagrange interpolatory polynomials $H_{n1}^{(\alpha,\beta)}(f, x) = \sum_{k=1}^{n} f(x_k) l_k^{(\alpha,\beta)}(x)$, whenever m is *odd* supposing certain conditions on coefficients e_{ik} while the cases of even values of m are similar to the Hermite–Fejér interpolation (cf. [1; Parts 3 and 4], [2, Theorems 2.1 and 2.7]).

The accomplishment of the corresponding conditions on e_{ik} for odd m was investigated in [1] and R. Sakai [4]; a (sometimes sketchy) proof was given in [4] if $\alpha = \beta = -1/2$. While the ideas were nice and correct, in many places a more detailed, finer and sometimes a modified argument may be useful. This plan is realized by this paper for arbitrary fixed $\alpha, \beta > 1$.

2. Notations and preliminary results

For arbitrary X and $m \geqq 1$ we have

$$e_{0k} = 1 \tag{2.1}$$

$$e_{tk} = -\frac{1}{t!} \sum_{r=0}^{t-1} (t)_r e_{rk} \{l_k^m(x)\}_{x=x_k}^{(t-r)}, \quad 1 \leqq t \leqq m-1, \quad (m \geqq 2) \tag{2.2}$$

where $(t)_0 = 1$, $(t)_r = t(t-1)\ldots(t-r+1)$, $r > 0$,

$$l_k^{(r)}(x_k) = \frac{\omega_n^{(r+1)}(x_k)}{(r+1)w_n'(x_k)} \tag{2.3}$$

for arbitrary k, $1 \leqq k \leqq n$ (cf. [2, (3.2), (3.4) and (3.8)]).

From now on let $\omega_n(x) = P_n^{(\alpha,\beta)}(x)$, $\alpha, \beta > -1$ are fixed, with

$$P_n(1) = \binom{n+\alpha}{n} \sim n^\alpha, \quad P_n^{(\alpha,\beta)}(-x) = (-1)^n P_n^{(\beta,\alpha)}(x) \tag{2.4}$$

($a_n \sim b_n$ iff $0 < c_1 \leqq a_n/b_n \leqq c_2$, $b_n \neq 0$).

The following facts are well known. If $x = \cos\vartheta$, $x_{kn}^{(\alpha,\beta)} = \cos\vartheta_{kn}^{(\alpha,\beta)}$, $1 \leqq k \leqq n$, are the roots of $P_n^{(\alpha,\beta)}(x)$ then with $x_{0n} = 1$, $x_{n+1,n} = -1$, $\vartheta_{0n} = 0$, $\vartheta_{n+1,n} = \pi$,

$$\vartheta_{k+1} - \vartheta_k \sim \frac{1}{n}, \quad \vartheta_k \sim \frac{k}{n}, \quad k = 0, 1, \ldots, n. \tag{2.5}$$

Further if x_j is the nearest root(s) to x ($j = j(n)$),

$$|x - x_k| \sim \frac{|j-k|(j+k)}{n^2}, \quad k \neq j,$$
$$|x - x_j| \leqq c\frac{j}{n^2}, \tag{2.6}$$

$$|P_n(x)| \sim |x - x_j|\frac{n^{\alpha+2}}{j^{\alpha+3/2}} \sim |\vartheta - \vartheta_j|\frac{n^{\alpha+1}}{j^{\alpha+1/2}} \leqq c\frac{n^\alpha}{j^{\alpha+1/2}} \tag{2.7}$$

uniformly in $x \in [-1+\varepsilon, 1]$. Finally we need

$$|P_n'(x_k)| \sim k^{-\alpha-3/2}n^{\alpha+2}, \qquad 0 < \vartheta_k \leqq \pi - \varepsilon.$$

(See [3; (4.1.1), (4.1.3), (8.9.2)] and [2] for other references; the symbol "$\sim$" is uniform in n and k (cf. [3; 1.1]); $c, c_1 \ldots$ may denote different constants; their dependence of certain parameters will be clear from the corresponding formulae; $0 < \varepsilon < 2$, fixed.)

Using the differential equation

$$(1-x^2)P_n^{(s)}(x) + \{\beta - \alpha - (\alpha+\beta+2s-2)x\}P_n^{(s-1)}(x) +$$
$$+\{n(n+\alpha+\beta+1) - (s-2)(\alpha+\beta+s-1)\}P_n^{(s-2)}(x) = 0, \quad s = 2, 3, \ldots \tag{2.8}$$

([2; (3.16)]) we get the following statement. If $K = \min(k, n-k+1)$, $d(j,k) = (\alpha - \beta + (\alpha+\beta+2j)x_k)/2$ and $M \geqq 1$, fixed integer, then we have

STATEMENT 2.1. *Let $1 \leqq r \leqq M$. Then we can write*

$$l_k^{(r)}(x_k) = \begin{cases} \dfrac{(-1)^j}{2j+1}\left(\dfrac{n}{\sin\vartheta_k}\right)^{2j}(1+\varepsilon_k) & \text{if } r = 2j \\ \dfrac{(-1)^j}{\sin^2\vartheta_k}\left(\dfrac{n}{\sin\vartheta_k}\right)^{2j}\{d(j+1,k)+\varepsilon_k\} & \text{if } r = 2j+1 \end{cases} \tag{2.9}$$

for $1 \leqq k \leqq n$, $n \geqq 2$.

Here and later $\{\varepsilon_k\} = \{\varepsilon_{kn}^{(\alpha,\beta)}(r)\}$ denotes a properly given sequence which may be different even in subsequent formulae. However, they fulfil

$$|\varepsilon_{kn}^{(\alpha,\beta)}(r)| \leqq c(\alpha,\beta,M)\left(\frac{1}{n} + \frac{1}{K^2}\right), \quad 1 \leqq r \leqq M, \quad 1 \leqq k \leqq n, \quad n \geqq 2,$$

where $c(\alpha, \beta, M) > 0$ does not depend on its occurrence.

Statement 2.1 can be derived from [2, Lemma 3], but we will prove it (Part 4), for sake of completeness.

3. Results

By previous notations and (2.9) we will prove

THEOREM 3.1. *For arbitrary real s with $|s| \leqq M$ we can write if $1 \leqq k \leqq n$, $n = 1, 2, \ldots$*

$$(3.1) \qquad (l_k^s)^{(2j)} := \left\{ (l_{kn}^{(\alpha,\beta)}(x))^s \right\}_{x=x_k}^{(2j)} = (-1)^j \left(\frac{n}{\sin \vartheta_k} \right)^{2j} \{ p_j(s) + q_{2j}(s) \}$$

if $j = 0, 1, 2, \ldots, [(M-1)/2]$ where

$$(3.2) \qquad \begin{cases} p_0(s) = 1, \quad p_j(s) = \sum_{i=1}^{j} (-1)^{j-i} a(i,j) s^i, \quad j \geqq 1 \\ q_0(s) = 0, \quad q_{2j} \in \mathcal{P}_{2j}, \quad q_{2j} \text{ depends on } k \text{ and } n \end{cases}$$

further they fulfil the relations

$$(3.3) \qquad p_j(s+1) = \frac{1}{2j+1} \sum_{i=0}^{j} \binom{2j+1}{2i} p_i(s), \quad j \geqq 1,$$

$$(3.4) \qquad \left| \frac{d^t q_{2j}(s)}{ds^t} \right| \leqq |\varepsilon_k|, \quad t = 0, 1, \ldots, j.$$

By (3.2) and (3.3), the coefficients $a(i,j)$ do not depend on k or n, even they are independent of α and β. Further it is easy to get

$$(3.5) \qquad \begin{cases} p_j(0) = 0 \quad (j \geqq 1), \quad p_j(1) = \dfrac{1}{2j+1} \\ p_j(t) > 0 \quad \text{for } t = 1, 2, \ldots, M. \end{cases}$$

Again by (3.2)–(3.3) one can successively get

$$p_0(s) = 1, \quad p_1(s) = \frac{s}{3}, \quad p_2(s) = \frac{5s^2 - 2s}{15},$$

$$p_3(s) = \frac{35s^3 - 42s^2 + 16s}{63}, \quad p_4(s) = \frac{175s^4 - 420s^3 + 404s^2 - 144s}{135}, \ldots$$

etc. (cf. [1] or [4]) which suggest

THEOREM 3.2. *With the previous conditions and notations*

$$a(i,j)>0, \quad i=1,2,\dots,j, \quad j\geqq 1. \tag{3.6}$$

An important consequence of (3.6) is that

$$p_j(t)\neq 0 \quad \text{if} \quad |t|=1,2,\dots,M, \quad j\geqq 1. \tag{3.7}$$

Indeed, if $u>0$, we remarked it before; if $u=-t$, $t>0$, by definition $p_j(-t)=(-1)^j\sum_{i=1}^{j}a(i,j)t^i$ whence (3.6) gives (3.7).

By (3.7) and (3.4)

$$(-1)^j(l_k^t)^{(2j)}\sim\left(\frac{n}{\sin\vartheta_k}\right)^{2j} \quad \text{if} \quad K\geqq c_0, \quad |t|=1,2,\dots,M \tag{3.8}$$

(cf. [2; (3.30)]).

DEFINITION. Let $R_0(s)=1$ and

$$R_j(s)=\sum_{i=1}^{j}a(i,j)s^i, \quad j\geqq 1, \quad |s|\leqq M. \tag{3.9}$$

Relations (3.9), (3.7) and (3.6) immediately give

$$p_j(-s)=(-1)^jR_j(s), \qquad R_j(-s)=(-1)^jp_j(s), \quad j\geqq 0$$

$$R_j(s)>0 \quad \text{if} \quad s>0, \quad R_j(t)\neq 0 \quad \text{if} \quad |t|=0,1,\dots,M, \quad j\geqq 1.$$

Now we can formulate our main relations.

THEOREM 3.3. *For arbitrary fixed $m\geqq 1$ and $\alpha,\beta>-1$ we have for $1\leqq k\leqq n$, $n=1,2,\dots$*

$$\begin{cases} e_{0knm}=1, \\ e_{2t,knm}=\frac{1}{(2t)!}R_t(m)\left(\frac{n}{\sin\vartheta_k}\right)^{2t}(1+\varepsilon_k), \quad 1\leqq t\leqq\left[\frac{m-1}{2}\right]. \end{cases} \tag{3.10}$$

By (3.10)

$$e_{2t,k}\sim\left(\frac{n}{\sin\vartheta_k}\right)^{2t}, \quad 1\leqq t\leqq\left[\frac{m-1}{2}\right] \quad \text{if} \quad K\geqq c_0. \tag{3.11}$$

An important application: If $m=1,3,\dots$ is fixed, odd, then for a proper $f\in C$

$$\overline{\lim_{n\to\infty}}\|H_{nm}^{(\alpha,\beta)}(f,x)\|_C=\infty, \quad \alpha,\beta>-1, \quad \text{fixed}$$

(cf. [1; Th. 3], [2; Th. 2.7] and [4]).

4. Proofs

PROOF OF STATEMENT 2.1. We use induction. (2.8) with $s=2$ gives

$$\frac{P_n''}{P_n'} = \frac{\alpha-\beta+(\alpha+\beta+2)x_k}{\sin^2\vartheta_k},$$

whence by (2.3) we get (2.9) for $l_k'(x_k)$ (with $\varepsilon_k(1)=0$). If $s=3$, from (2.8), by the previous formula

$$\frac{P_n'''}{P_n'} = -\frac{n^2}{\sin^2\vartheta_k}\left\{1+\frac{\alpha+\beta+1}{n}-\frac{\alpha+\beta+2}{n^2}-\right.$$
$$\left.-\frac{\alpha-\beta+(\alpha+\beta+4)x_k}{n^2}\,\frac{\alpha-\beta+(\alpha+\beta+2)x_k}{\sin^2\vartheta_k}\right\}$$

which proves (2.9) for $l_k''(x_k)$ with $\varepsilon_k(2)=\{\dots\}-1$ (cf. (2.3); we use $(n\sin\vartheta_k)^{-2}\leqq cK^{-2}$ (see 2.5)).

Now supposing (2.9) for $r=1,2,\dots,t$ we get by (2.3) and (2.8)

$$(t+2)l_k^{(t+1)}(x_k) = \frac{P_n^{(t+2)}}{P_n'} =$$
$$= \frac{\alpha-\beta+(\alpha+\beta+2t+2)x_k}{\sin^2\vartheta_k}\,\frac{P_n^{(t+1)}}{P_n'}-$$
$$-\frac{n^2}{\sin^2\vartheta_k}\left\{1+\frac{\alpha+\beta+1}{n}-\frac{t(\alpha+\beta+t+1)}{n^2}\right\}\frac{P_n^{(t)}}{P_n'}$$

whence we get (2.9) for $r=t+1$ (cf. [2, Lemma 3.4] for other details). □

PROOF OF THEOREM 3.1. When $j=0$, (3.1) obviously holds. If $j\geqq 1$ we define the polynomials p_j and q_{2j} as follows. For arbitrary real s

(4.1)
$$(l_k^s(x))^{(J)} =$$
$$= \sum_{\substack{i_1+i_2+\dots+i_t=J\\ 1\leqq i_1\leqq i_2\leqq\dots\leqq i_t\leqq J}} A(I)(s)_t l_k^{s-t}(x)l_k^{(i_1)}(x)l_k^{(i_2)}(x)\dots l_k^{(i_t)}(x),\quad J\geqq 1,$$

where $A(I)=A(i_1,i_2,\dots,i_t)>0$, integer. E.g.,

(4.2)
$$(l_k^s(x))' = sl_k^{s-1}(x)l_k'(x),\ (l_k^s(x))'' = (s)_2 l_k^{s-2}(x)(l_k'(x))^2+\underline{sl_k^{s-1}(x)l_k''(x)},$$
$$(l_k^s(x))''' = (s)_3 l_k^{s-3}(x)(l_k'(x))^3+3(s)_2 l_k^{s-2}(x)l_k'(x)l_k''(x)+sl_k^{s-1}(x)l_k'''(x),$$
$$(l_k^s(x))^{(\mathrm{IV})} = (s)_4 l_k^{s-4}(x)(l_k'(x))^4+6(s)_3 l_k^{s-3}(l_k'(x))^2 l_k''(x)+$$
$$+\underline{3(s)_2 l_k^{s-2}(x)(l_k''(x))^2}+4(s)_2 l_k^{s-2}(x)l_k'(x)l_k'''(x)+\underline{sl_k^{s-1}(x)l_k^{(\mathrm{IV})}(x)}.$$

Now let $J = 2j$ and $x = x_k$. Consider a term for which all the i_r are even. Then by $i_r \geqq 2$ and $\sum_{r=1}^{t} i_r = 2j$, $t \leqq j$, by (2.9) its sign is $(-1)^{\sum_{r=1}^{t} i_r/2} = (-1)^j$ and its exact order is $(n/\sin\vartheta_k)^{\sum_{r=1}^{t} i_r} = (n/\sin\vartheta_k)^{2j}$. If $\sum^{(1)}$ denotes the sum of these terms (see the underlinings at (4.2)), we write

$$\sum^{(1)}\Big\{A(I)(s)_t \prod_{r=1}^{t} l_k^{(i_r)}\Big\} =$$

$$= (-1)^j\Big(\frac{n}{\sin\vartheta_k}\Big)^{2j} \sum^{(1)}\Big\{A(I)(s)_t \prod_{r=1}^{t} \frac{1}{i_r+1}(1+\varepsilon_k(i_r))\Big\} :=$$

$$:= (-1)^j\Big(\frac{n}{\sin\vartheta_k}\Big)^{2j}\Big\{\sum^{(1)}\Big(A(I)(s)_t \prod_{r=1}^{t} \frac{1}{i_r+1}\Big) + u_j(s)\Big\}$$

where $u_j \in \mathcal{P}_j$ and as it is easy to get, $\left|\frac{d^\nu u_j}{ds^\nu}\right| \leqq |\varepsilon_k|$, $0 \leqq \nu \leqq j$. Let $p_j(s) := \sum^{(1)}\Big(A(I)(s)_t \prod_{r=1}^{t} \frac{1}{i_r+1}\Big)$. Now take a term in which there are (at least two) odd i_r. Using (2.9) it can be estimated by $\sin^{-4}\vartheta_k (n/\sin\vartheta_k)^{2j-2} = (n\sin\vartheta_k)^{-2}(n/\sin\vartheta_k)^{2j} \leqq |\varepsilon_k|(n/\sin\vartheta_k)^{2j}$ (by $(n\sin\vartheta_k)^{-2} \leqq cK^{-2}$) whence if their sum is $\sum^{(2)}$ for $v_{2j}(s) := \sum^{(2)}\Big(A(I)(s)_t \prod_{r=1}^{t} l_k^{(i_r)}\Big)$ we obviously have $\left|\frac{d^\nu v_{2j}}{ds^\nu}\right| \leqq |\varepsilon_k|\left(\frac{n}{\sin\vartheta_k}\right)^{2j}$, $0 \leqq \nu \leqq j$, i.e. we obtain (3.4) with $q_{2j} := u_j + (-1)^j(\sin\vartheta_k/n)^{2j} v_{2j}$. For example if $j = 1$, we have (cf. (4.2) and (2.9))

$$(l_k^s)'' = -\frac{s}{3}\Big(\frac{n}{\sin\vartheta_k}\Big)^2(1+\varepsilon_k(2)) + s(s-1)\frac{(d(1,k)+\varepsilon_k(1))^2}{\sin^4\vartheta_k} =$$

$$= -\Big(\frac{n}{\sin\vartheta_k}\Big)^2\Big\{\frac{s}{3} + \varepsilon_k(2)\frac{s}{3} + \frac{(d(1,k)+\varepsilon_k(1))^2}{n^2\sin^2\vartheta_k}s(s-1)\Big\},$$

i.e. $p_1(s) = \frac{s}{3}$ and $q_k(s) = \{\ldots\} - \frac{s}{3}$ (cf. the proof of [2; Lemma 3.9] for a more detailed argument).

These definitions give (3.1), (3.2) and (3.4). To prove (3.3) we write by Leibniz rule

$$(l_k^{s+1})^{(2j)} = (l_k^s l_k)^{(2j)} = \sum_{i=0}^{2j}\binom{2j}{i}(l_k^s)^{(i)} l_k^{(2j-i)} = \tag{4.4}$$

$$= \sum_{i=0}^{j} \binom{2j}{2i} (l_k^s)^{(2i)} l_k^{(2j-2i)} + \sum_{i=1}^{j} \binom{2j}{2i-1} (l_k^s)^{(2i-1)} l_k^{(2j-2i+1)} := S_1 + S_2.$$

Here $S_2 \in \mathcal{P}_{2j}$ and it can be handled as the sum $\sum^{(2)}$ before. For S_1, using that $p_j(1) = 1/(2j+1)$ (use (2.9) and the definition of $p_j(s)$) we write, using induction and previous considerations,

$$S_1 = (-1)^j \Big(\frac{n}{\sin \vartheta_k}\Big)^{2j} \sum_{i=0}^{j} \binom{2j}{2i} (p_i(s) + q_{2i}(s)) \Big(\frac{1}{2j-2i+1} + q_{2j-2i}(1)\Big) =$$

$$= (-1)^j \Big(\frac{n}{\sin \vartheta_k}\Big)^{2j} \Big(\sum_{i=0}^{j} \binom{2j}{2i} \frac{p_i(s)}{2j-2i+1} + u_{2j}(s)\Big)$$

where $u_{2j} \in \mathcal{P}_{2j}$ and $|u_{2j}^{(t)}| \leqq |\varepsilon_k|$, $0 \leqq t \leqq j$. Noting

$$\binom{2j}{2i} \frac{1}{2j-2i+1} = \binom{2j+1}{2i} \frac{1}{2j+1},$$

we obtain (3.3) □

PROOF OF THEOREM 3.2. If

$$D = D_{kn} := \{(x, s); |x - x_k| < \delta_n \text{ and } |s| < M_1\}$$

(k and n are fixed $\delta_n > 0$ is small enough) then, obviously, $L(x, s) := l_{kn}^s(x)$ is differentiable finitely many times (as a function of two variables) whenever $(x, s) \in D$.

LEMMA 4.1. *If $0 \leqq u, v \leqq M_2$ and $\delta_n > 0$ is small enough, then*

$$\Big(\frac{d^v}{ds^v}\Big(\frac{\partial^u L}{\partial x^u}\Big)_{x=x_0}\Big)_{s=s_0} = \Big(\frac{d^u}{dx^u}\Big(\frac{\partial^v L}{\partial s^v}\Big)_{s=s_0}\Big)_{x=x_0}, \quad (x_0, s_0) \in D. \tag{4.5}$$

Indeed, applying $\partial^2 F/(\partial x \partial s) = \partial^2 F/(\partial s \partial x)$ (Young's theorem) for L and its partial derivatives we get that either side of (4.5) is equal to $(\partial^{u+v} L/(\partial x^u \partial s^v))_{x=x_0, s=s_0}$. □

By (3.1), (3.2) and (3.4)

$$\Big(\frac{d^t}{ds^t}\Big(\frac{\partial^{2j} L}{\partial x^{2j}}\Big)_{x=x_k}\Big)_{s=0} = (-1)^j \Big(\frac{n}{\sin \vartheta_k}\Big)^{2j} \{p_j^{(t)}(s) + q_{2j}^{(t)}(s)\}_{s=0} =$$

$$= \Big(\frac{n}{\sin \vartheta_k}\Big)^{2j} \{(-1)^t t! a(t, j) + \varepsilon_k\}. \tag{4.6}$$

Changing the order of derivatives (cf. [4])) we get, using $\partial^t L/\partial s^t = l_k^s(x)\log^t l_k(x)$,

$$\left(\frac{d^{2j}}{dx^{2j}}\left(\frac{\partial^t L}{\partial s^t}\right)_{s=0}\right)_{x=x_k} = (\log^t l_k(x))^{(2j)}_{x=x_k}. \tag{4.7}$$

Using

$$\begin{aligned}(\log l_k(x))^{(2j)} &= \left(\frac{l_k'(x)}{l_k(x)}\right)^{(2j-1)} = \left(\sum_{i\neq k}\frac{1}{x-x_i}\right)^{(2j-1)} = \\ &= -(2j-1)!\sum_{i\neq k}\frac{1}{(x-x_i)^{2j}},\end{aligned} \tag{4.8}$$

by (4.5)–(4.8) we get

$$0 < (-1)^t(\log^t l_k(x))^{(2j)}_{x=x_k} = \left(\frac{n}{\sin\vartheta_k}\right)^{2j}\{t!a(t,j)+\varepsilon_k\}, \quad j\geqq 1, \tag{4.9}$$

when $t=1$, whence $a(1,j)>0$. To obtain (4.9) for $t>1$, we prove

$$\left|(\log^t l_k(x))^{(2j-1)}_{x=x_k}\right| \leqq |\varepsilon_k|\left(\frac{n}{\sin\vartheta_k}\right)^{2j}, \quad j\geqq 1, \tag{4.10}$$

first when $t=1$.

Indeed, by induction for j, $(\log l_k(x))^{(2j-1)}_{x=x_k} = \sum B(I)l_k^{(i_1)}l_k^{(i_2)}\dots l_k^{(i_r)}$ with certain coefficients $B(I)=B(i_1,i_2,\dots,i_T)$, where $\sum i_r = 2j-1$, $1\leqq i_1\leqq i_2\leqq \dots i_T\leqq 2j-1$. But then at least one i_r is odd, whence by (2.9) we get (4.10) as before. Now we prove (4.9) when $t=2$ (whence by (4.5)–(4.8), $a(2,j)>0$). Indeed, by

$$\begin{aligned}(\log^2 l_k(x))^{(2j)} &= (\log l_k(x)\cdot\log l_k(x))^{(2j)} = \\ &= \sum_{i=0}^{j}\binom{2j}{2i}(\log l_k(x))^{(2i)}(\log l_k(x))^{(2j-2i)} + \\ &+ \sum_{i=1}^{j}\binom{2j}{2i-1}(\log l_k(x))^{(2i-1)}(\log l_k(x))^{(2j-2i+1)}.\end{aligned}$$

Using now (4.9) and (4.10) for $t=1$, we get (4.9) when $t=2$. The further steps are obvious. □

Proof of Theorem 3.3. Let

$$C_j(s) := \sum_{i=0}^{j}(-1)^i\binom{2j}{2i}R_i(s)p_{j-i}(s), \quad j\geqq 1,\ |s|\leqq M. \tag{4.11}$$

The main ingredient of the proof is the relation

$$C_j(t) = 0, \quad j \geqq 1, \quad t = 0, \pm 1, \ldots, M \tag{4.12}$$

(cf. [1; Th. 4]). $C_j(0) = 0$ is obvious by $p_j(0) = R_j(0) = 0$ $(j \geqq 1)$. Now we remark that $C_j(-s) = C_j(s)$ for any real $|s| \leqq M$. Indeed, by $p_i(-s) = (-1)^i R_i(s)$ and $R_i(-s) = (-1)^i p_i(s)$, $i \geqq 0$,

$$C_j(-s) = \sum_{i=0}^{j} (-1)^i \binom{2j}{2i} R_i(-s) p_{j-i}(-s) =$$

$$= \sum_{i=0}^{j} (-1)^{j-i} \binom{2j}{2i} p_i(s) R_{j-i}(s) = C_j(s),$$

as it was stated. This means it is enough to verify (4.12) only for $t = 1, 2, \ldots, M$. By

$$0 = (l_k^{-1+1})^{(2j)} = \sum_{i=0}^{2j} \binom{2j}{i} (1/l_k)^{(i)} (l_k)^{(2j-i)} \quad (j \geqq 1)$$

we get as in Part 4.1

$$0 = p_j(0) = \sum_{i=0}^{j} \binom{2j}{2i} p_i(-1) p_{j-i}(1) =$$

$$= \sum_{i=0}^{j} (-1)^i \binom{2j}{2i} R_i(1) p_{j-i}(1) = C_j(1).$$

Now, induction. Supposing $C_j(t) = C_j(-t) = 0$, by (4.11) and (3.3)

$$0 = C_j(-t) =$$

$$= \sum_{i=0}^{j} (-1)^i \binom{2j}{2i} R_i(-t) \left\{ \frac{1}{2j-2i+1} \sum_{k=0}^{j-i} \binom{2j-2i+1}{2k} p_k(-t-1) \right\} := I.$$

Here using the relation

$$\binom{2j}{2i} \frac{1}{2j-2i+1} \binom{2j-2i+1}{2k} = \binom{2j}{2k} \frac{1}{2j-2k+1} \binom{2j-2k+1}{2i}$$

and changing the order of summations

$$I = \sum_{k=0}^{j} (-1)^k \binom{2j}{2k} R_k(t+1) \left\{ \sum_{i=0}^{j-k} \frac{1}{2j-2k+1} \binom{2j-2k+1}{2i} p_i(t) \right\} = C_j(t+1)$$

which completes the proof of (4.12).

Relations (3.10) is proved by induction, again. If $t = 0$, (3.10) holds true. Supposing it until $t - 1$, we write by (2.2)

$$e_{2t,k} = -\frac{1}{(2t)!}\Big\{\sum_{i=0}^{t-1}(2t)_{2i}e_{2i,k}(l_k^m)^{(2t-2i)} + \sum_{i=0}^{t-1}(2t)_{2i+1}e_{2i+1,k}(l_k^m)^{(2t-2i-1)}\Big\}.$$

Here the second sum can be estimated by $|\varepsilon_k|(n/\sin\vartheta_k)^{2t}$ (see [2, Lemma 3.11]). For the first sum, J, by (3.10) and Theorem 3.1

$$J = \frac{1+\varepsilon_k}{(2t)!}\Big(\frac{n}{\sin\vartheta_k}\Big)^{2t}\Big\{(-1)^{t+1}\sum_{i=0}^{t-1}(-1)^i\binom{2t}{2i}R_i(m)p_{t-i}(m)\Big\}$$

which gives (3.10) for t, considering that by (4.12), $\{\ldots\} = R_t(m)$. □

Addendum. Recent results corresponding to $e_{2t+1,knm}$ (see Theorem 3.3) are in our subsequent paper [5].

REFERENCES

[1] SAKAI, R., Hermite–Fejér interpolation prescribing higher order derivatives, *Progress in approximation theory*, Academic Press, Boston, MA, 1991, 731–759. *MR* **92m**: 41017

[2] VÉRTESI, P., Hermite–Fejér interpolations of higher order. I, *Acta Math. Hungar.* **54** (1989), 135–152. *MR* **90k**: 41008

[3] SZEGŐ, G., *Orthogonal Polynomials*, Third edition, American Mathematical Society Colloquium Publications, Vol. 23, American Mathematical Society, Providence, R. I., 1967. *MR* **46** #9631

[4] SAKAI, R., Certain unbounded Hermite–Fejér interpolatory polynomial operators, *Acta Math. Hungar.* **59** (1992), 111–114.

[5] SAKAI, R. and VÉRTESI, P., Hermite–Fejér interpolations of higher order. IV, *Studia Sci. Math. Hungar.* **28** (1993) (to appear).

(Received May 2, 1990)

DEPARTMENT OF MATHEMATICS
KARIYA-HIGASHI SENIOR HIGH SCHOOL
HAJODO-CHO MITSUMATA 20
AICHI KARIYA 448
JAPAN

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 99–104

LOWER BOUNDS FOR THE NUMBERS OF EXTREMAL AND EXPOSED DIAMETERS OF A CONVEX BODY

V. P. SOLTAN and M. H. NGUEN

1. Introduction

Further A will denote a convex body (i.e. a compact convex set with non-empty interior) in n-dimensional euclidean space E^n. A chord $[a, b]$ of A is said to be an *affine diameter* of A if and only if there exists a pair of different parallel hyperplanes each containing one of the endpoints a, b and supporting A. It is known that every point $x \in A$ belongs to at least one affine diameter of A, and for any direction in E^n, there is at least one affine diameter of A parallel to this direction [1].

An affine diameter $[a, b]$ will be called *extremal* for A if $a, b \in \operatorname{ext} A$, and will be called *exposed* for A if $A \cap H = \{a\}$, $A \cap G = \{b\}$ for some parallel hyperplanes H, G supporting A. Any convex body in E^n has no more than 2^n points each being the endpoints of an extremal diameter, and has exactly 2^n such points if and only if it is an n-dimensional parallelepiped [2]. Up to now the following problem is not solved: to determine the least upper bound for the cardinality c of points each two being the endpoints of an exposed diameter of A. The equality $c = 2n - 1$ was proposed in [3], but from [4] it follows that $c \geqq (1.15)^n$.

2. Main theorems

Denote by $p(A)$ and $q(A)$ the numbers of all extremal and exposed diameters of A, respectively.

THEOREM 1. *The numbers $p(A)$ and $q(A)$ are positive. Any of them is finite if and only if A is a polytope.*

THEOREM 2. $p(A) \geqq n(n+1)/2$. *Equality $p(A) = n(n+1)/2$ is fulfilled if and only if A is an n-dimensional simplex.*

THEOREM 3. $q(A) \geqq n$. *Equality $q(A) = n$ is fulfilled if and only if A is an n-dimensional octahedron.*

1980 *Mathematics Subject Classification* (1985 Revision). Primary 52A20; Secondary 52A37.

Key words and phrases. Convex sets in n dimensions, inequalities, extremum problems.

Akadémiai Kiadó, Budapest

3. Some auxiliary lemmas

LEMMA 1. *Each point $x \in \operatorname{ext} A$ is the end of some extremal diameter of A.*

PROOF. Let L be some hyperplane supporting A at x, and denote by N another hyperplane parallel to L and supporting A. If y is any extremal point of $A \cap N$, then $y \in \operatorname{ext} A$. Hence $[x, y]$ is an extremal diameter of A. □

Recall that a segment $[a, b] \subset A$ is named a metric diameter of A if $\|a - b\| = \max\{\|u - v\| : u, v \in A\}$.

LEMMA 2. *Any metric diameter $[a, b]$ of A is an exposed diameter of A.*

PROOF. Let L, N be two hyperplanes passing through a, b perpendicularly to segment $[a, b]$. It is well-known that L, N support A, and $A \cap L = \{a\}$, $A \cap N = \{b\}$. This means that $[a, b]$ is an exposed diameter of A. □

LEMMA 3. *Any vertex a of a convex polytope $B \subset E^n$ is the end of some exposed diameter of B.*

PROOF. Let H be a hyperplane satisfying the condition $A \cap H = \{a\}$. Denote by G another hyperplane parallel to H and supporting B. We can slightly move hyperplanes H and G in order to make the set $B \cap G$ one-vertex and to preserve the relation $B \cap H = \{a\}$. If $B \cap G = \{b\}$, then $[a, b]$ is an exposed diameter of A. □

A point $a \in A$ is named k-extremal for A if it does not belong to the relative interior of any $(k+1)$-dimensional simplex $\delta \subset A$ and belongs to the relative interior of some k-dimensional simplex $\sigma \subset A$ [5]. In our notations, ext A consists of all 0-extremal points.

LEMMA 4. *Let A be a union of some convex bodies $B_0, B_1, \ldots, B_m$ whose interiors are pairwise disjoint. If a point $x \in \operatorname{int} A$ is k-extremal for B_0, then it belongs to at least $n - k$ of the sets $B_1, \ldots, B_m$.*

Proof of Lemma 4 will be organized by induction on n. The case $n = 1$ is trivial. Suppose the assertion of Lemma 4 is true for all $n \leqq l - 1$, and let A be a convex body in E^l.

First the case $k > 0$ will be considered. Let $\delta \subset B_0$ be any k-dimensional simplex containing x in its relative interior, and H be some hyperplane satisfying the conditions:

$$x \in H, \quad \delta \subset H, \quad \operatorname{int} B_0 \cap H \neq \emptyset.$$

Let us denote by $C_1, \ldots, C_p$ all the members of the family $\{B_1, \ldots, B_m\}$ whose interiors intersect H. Relative to H, a convex body $A \cap H$ is the union of convex bodies $B_0 \cap H, C_1 \cap H, \ldots, C_p \cap H$ having pairwise disjoint interiors. By the inductive hypothesis, x belongs to at least $(n-1) - (k-1) = n - k$

members of the family $\{C_1 \cap H, \dots, C_p \cap H\}$. Hence x belongs to at least $n-k$ sets from $\{B_1, \dots, B_m\}$.

Now let us consider the case $k=0$. Without loss of generality we may suppose that x belongs to the sets $B_0, B_1, \dots, B_q$, $q \leqq m$, and only to them. Then, for some $r>0$, the ball $\Sigma_r(x)$ of radius r and centre x is contained in int A and does not intersect each of the sets $B_{q+1}, \dots, B_m$. Therefore from the relation $A = B_0 \cup B_1 \cup \dots \cup B_m$ one has

$$\Sigma_r(x) = [B_0 \cap \Sigma_r(x)] \cup \dots \cup [B_q \cap \Sigma_r(x)].$$

We are going to show that B_0 is locally conic at x. Indeed, let y be any point from bd $B_0 \cap \Sigma_r(x)$. Then $y \in B_i$ for some index $i = 1, \dots, q$. Because of convexity of B_0 and B_i, we obtain the inclusion $[x, y] \subset B_0 \cap B_i$. Hence $[x, y] \subset$ bd B_0. The last inclusion means that B_0 is locally conic at x. In this case, from the relation

$$A = \operatorname{conv}(\{x\} \cup (\operatorname{ext} A \setminus \{x\}))$$

it follows the existence of a one-dimensional face of B_0 of the form $[x, z]$. Any point $v \in]x, z[$ is 1-extremal for B_0. By the demonstrated above, v belongs to some $n-1$ cones $D_1, \dots, D_{n-1}$ from the family $\{B_1, \dots, B_q\}$. Hence $q \geqq n$.

Denote by L any two-dimensional plane passing through $[x, z]$ and intersecting int A. A convex figure $C = B_0 \cap L$ is locally conic at x. Therefore two segments of the form $[x, z]$, $[x, s]$ exist in bd C. Let us select any point $w \in]x, s[\cap \Sigma_r(x)$. Since $]v, w[\subset$ int B_0 and $B_1, \dots, B_m$ are convex, one has $w \in B_j$ for some set B_j different from $D_1, \dots, D_{n-1}$. Hence $q \geqq n$. □

Let $\mathcal{K}$ be any family of cones in E^n with common zero vertex. Two cones $L, M \in \mathcal{K}$ will be called *antipodal* provided $L \cap (-M) \neq \{0\}$ (or $M \cap (-L) \neq \{0\}$, which is the same). Denote by $t(\mathcal{K})$ the number of all pairs of antipodal cones in $\mathcal{K}$.

LEMMA 5. *Let $\mathcal{K} = \{K_1, \dots, K_m\}$ be a family of closed convex acute cones with common zero vertex satisfying the conditions*

$$\bigcup_{i=1}^{n} K_i = E^n, \quad \operatorname{int} K_i \cap \operatorname{int} K_j = \emptyset, \quad i \neq j. \tag{1}$$

Then $t(\mathcal{K}) \geqq n(n+1)/2$. Equality $t(\mathcal{K}) = n(n+1)/2$ holds if and only if $m = n+1$.

Proof of Lemma 5 will be carried out by induction on $n = \dim E^n$. The case $n=1$ is trivial: $m=2$ and $t=2$.

Let us suppose that the assertion of Lemma 5 is true for all $n \leqq k-1$, and $\mathcal{K}$ be a family of cones in E^k satisfying conditions (1). Denote by l some extremal ray of K_1. By Lemma 4, l belongs to at least $k-1$ cones from the

family $\mathcal{K}\setminus\{K_1\}$. The ray $-l$ belongs to a new cone $K_i\in\mathcal{K}$. Therefore K_i is antipodal to at least k cones from $\mathcal{K}$.

Denote by H some hyperplane whose intersection with K_i is equal to $\{0\}$. We may suppose, that $\operatorname{int}K_j\cap H\neq\emptyset$ for each cone K_j satisfying the condition $K_j\cap H\neq\emptyset$ (otherwise one can slightly move H in order to obtain the desired property). If $L_1,\dots,L_r$ are the cones in $\mathcal{K}$ intersected by H, then in the $(k-1)$-dimensional space H the family of cones $\mathcal{L}=\{H\cap L_1,\dots,H\cap L_r\}$ satisfies conditions (1). By the inductive conjecture, we have $t(\mathcal{L})\geqq k(k-1)/2$. Hence the family $\{L_1,\dots,L_r\}$ contains at least $k(k-1)$ pairs of antipodal cones from $\mathcal{K}$. Taking into account k antipodal pairs containing K_i, we obtain

$$t(\mathcal{K})\geqq k(k-1)/2+k=k(k+1)/2.$$

If $m=k+1$, then every n cones in $\mathcal{K}$ have a common ray. Let l_i be a ray belonging to the intersection of cones K_j, $j\neq i$. Since $-l_i\notin\bigcup_{j\neq i}K_j$, one has $-l_i\subset K_i$. Hence each cone K_i belongs exactly to n pairs of antipodal cones from $\mathcal{K}$. From this observation, we obtain that $t(\mathcal{K})=n(n+1)/2$.

Conversely, let $t(\mathcal{K})=n(n+1)/2$, and suppose that $m>n+1$. Denote by K_i a cone from $\mathcal{K}$ having a maximum number s of antipodal cones in $\mathcal{K}$. Repeating previous considerations, we obtain inequality $t(\mathcal{K})\geqq n(n-1)/2+s$. Therefore $s=n$, and, by the inductive assumption, the hyperplane H intersects exactly n cones, say, $K_1,\dots,K_n$. From the condition $t(\mathcal{K})=n(n+1)/2$, we obtain that K_u is not contained in any pair of antipodal cones different from $\{K_i,K_u\}$. This situation can be realized only in the case $K_u\subset\operatorname{int}(-K_i)$. Since K_u has at least n faces of dimension $n-1$, there are at least n new cones in $\mathcal{K}$ intersecting $\operatorname{int}(-K_i)$. Thus $s>n$, which is in contradiction with the obtained relation $s=n$. Hence $m=n+1$. □

4. Proofs of main theorems

Proof of Theorem 1. From Lemma 1 it follows that $p(A)>0$. If A is a polytope with m vertices, then $p(A)\leqq\binom{m}{2}$. Conversely, if for a convex body $A\subset E^n$ the number $p(A)$ is finite, then the set $\operatorname{ext}A$ is finite, i.e. A is a polytope.

By Lemma 2, we have $q(A)>0$. If A is a polytope, then from the obvious inequality $q(A)\leqq p(A)$ it follows the finiteness of $q(A)$. Conversely, let A be a convex body having the finite number $q(A)=m$, and let $[c_i,b_i]$, $i=1,\dots,m$, be all the exposed diameters of A. Suppose A is not a polytope. Then A does not coincide with the polytope

$$B=\operatorname{conv}\{c_1,\dots,c_m,b_1,\dots,b_m\}.$$

Choose any point $z\in A\setminus B$ and a hyperplane H strictly separating z from B. Let R,S be two different hyperplanes parallel to H and supporting

A. Suppose that x lies between R and H. Denote by f a linear dilation of E^n with coefficient $r > 1$ in the direction orthogonal to H. If r is sufficiently big, then one of the ends of any metric diameter $[f(a), f(b)]$ of $f(A)$ does not lie between hyperplanes $f(H)$ and $f(S)$. By Lemma 2, $[f(a), f(b)]$ is an exposed diameter of $f(A)$. Therefore $[a, b]$ is an exposed diameter of A, and it is not contained in B. This fact is in contradiction with the choice of B. Hence A is a polytope. □

PROOF OF THEOREM 2. Following Theorem 1, we may suppose that A is a polytope. Let $a_1, \ldots, a_m$ be all the vertices of A. Let us denote by N_i the cone of all the external normals to A at a_i, and consider the cone $K_i = N_i - a_i$ obtained from N_i by a translation on vector $-a_i$. It is easy to see that $K_1, \ldots, K_m$ are closed convex polyhedral acute cones satisfying conditions (1). A chord $[a_i, a_j]$ is an extremal diameter of A if and only if the cones K_i, K_j are antipodal. Hence the assertion of Theorem 2 follows from Lemma 5. □

PROOF OF THEOREM 3. Following Theorem 1, we may suppose that A is a polytope. Let S be some $(n-1)$-dimensional face of A. Then S contains at least n vertices $a_1, \ldots, a_n$ of A. By Lemma 3, each point a_i is the end of some exposed diameter $[a_i, b_i]$. We can choose two parallel hyperplanes, which determine the exposedness of a_i, b_i, sufficiently close to aff S so that b_i will be situated out of S. In this case all the diameters $[a_i, b_i]$, $i = 1, \ldots, n$, are different. Hence $q(A) \geq n$.

If A is an n-dimensional octahedron, i.e.

$$A = \operatorname{conv}\{e_1, -e_1, \ldots, e_n, -e_n\}$$

for some linearly independent vectors $e_1, \ldots, e_n \in E^n$, then only the chords $[e_i, -e_i]$, $i = 1, \ldots, n$ are the exposed diameters of A. In this case $q(A) = n$.

Conversely, let $q(A) = n$ for some convex polytope A. By the demonstrated above, each $(n-1)$-dimensional face of A is a simplex with n vertices, and each vertex of A is the end of exactly one exposed diameter. Let S be any $(n-1)$-dimensional face of A. Denote by H a hyperplane parallel to aff S and supporting A out of S. Let $T = A \cap H$. We can take all the hyperplanes H_i, which determine the exposedness of S, sufficiently close to aff S such that parallel hyperplanes G_i support A at vertices of T. From these considerations it follows that T is an $(n-1)$-dimensional simplex (otherwise some vertex of T would be the end of at least two exposed diameters). Because of $q(A) = n$, all the exposed diameters of A are of the form $[a_i, b_i]$, $i = 1, \ldots, n$ where $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ are the vertices of S and T, respectively. From Lemma 3 it follows that $A = \operatorname{conv}(S \cup T)$.

Denote by K_i the convex cone with the apex a_i generated by S, and denote by N_i the convex cone with the apex b_i generated by T. We want to show that K_i and N_i are symmetric to each other relative to the point $c = (a_i + b_i)/2$. Suppose the contrary. Then the cones K_i and $(a_i - b_i) - N_i$

do not coincide. Let, for example, $K_i \subset (a_i - b_i) - N_i$ (the case $(a_i - b_i) - N_i \subset K_i$ is considered analogously). Then some vertex a_j, $j \neq i$, does not belong to $(a_i - b_i) - N_i$. In this case $[a_j, b_i]$ is a new exposed diameter of A, which is in contradiction with the above. Hence $K_i = (a_i - b_i) - N_i$.

The extremal rays of the cones K_i and N_i are generated by the edges $[a_i, a_j]$ and $[b_i, b_j]$, $j = 1, \ldots, n$, $j \neq i$, respectively. Hence the edges $[a_i, a_j]$ and $[b_i, b_j]$ are parallel to each other and the simplexes S, T are homothetic to each other with some coefficient $\mu < 0$.

Now we shall show that S and T are congruent. Suppose the contrary: let S be $|\mu|$ times ($\mu < -1$) greater than T. Then for any two vertices a_i, a_j from S, it is possible to choose parallel to each other hyperplanes P, Q supporting A at a_i, a_j, respectively, so that $A \setminus \{a_i, a_j\}$ lies strictly between P, Q. In this case $[a_i, a_j]$ is an exposed diameter of A. The last is impossible because of the assumption $q(A) = n$. Hence S and T are congruent.

From the above it follows that A is an octahedron. □

REFERENCES

[1] HAMMER, P. C., Diameters of convex bodies, *Proc. Amer. Math. Soc.* **5** (1954), 304–306. *MR* **15**–819

[2] DANZER, L. and GRÜNBAUM, B., Über zwei Probleme bezüglich konvexer Körper von P. Erdős und von V. L. Klee, *Math. Z.* **79** (1962), 95–99. *MR* **25** #1488

[3] GRÜNBAUM, B., Strictly antipodal sets, *Israel J. Math.* **1** (1963), 5–10. *MR* **28** #2480

[4] ERDÖS, P. and FÜREDI, Z., The greatest angle among n points in the d-dimensional Euclidean space, *Combinatorial mathematics* (Marseille-Luminy, 1981), North-Holland Math. Stud., 75, North-Holland, Amsterdam–New York, 1983, 275–283, *MR* **87g:** 52018. See also in the form: *Ann. Discrete Math.* **17** (1983), 275–283.

[5] ASPLUND, E., A k-extreme point is the limit of k-exposed points, *Israel J. Math.* **1** (1963), 161–162. *MR* **28** #4430

(Received June 8, 1989)

MATHEMATICAL INSTITUTE OF THE
ACADEMY OF SCIENCES
STR. ACADEMIEI 5
CHIŞINĂU
MOLDOVA

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 105–113

EXTENDING A QUASI-METRIC

J. DEÁK

Abstract

The non-symmetrical analogue of Hausdorff's theorem on extensions of compatible metrics holds for bounded quasi-metrics, but not for unbounded ones. Some related results on extending quasi-(pseudo)metrics will also be proved.

Hausdorff [16] proved that a compatible metric given on a closed subspace of a metrizable space has a compatible extension, see also [6, 21, 5]; the related question of continuously extending a pseudometric was investigated e.g. in [4, 12, 19, 14, 2, 1]. In this note we shall deal with analogous problems for quasi-(pseudo)metrics (see § 0 for definitions).

Extensions of pseudometrics can be of use when trying to extend uniformities [3, 15, 22]; in the non-symmetrical case we shall proceed conversely: results for quasi-(pseudo)metrics, at least for bounded ones, can be deduced from what is already known about extensions of quasi-uniformities [7 to 11], see § 1. The case of unbounded quasi-(pseudo)metrics is more delicate. The problem can be split into two: find extensions from dense subspaces (§ 3), respectively from closed ones (see § 2, which contains results for open subspaces, too).

§ 4 deals with similar questions in bitopological spaces.

§ 0. Preliminaries

0.1 Terminology. A non-negative real function d on $X \times X$ is a *quasi-pseudometric* on X ([23], see also [18, 13]) if $d(x,x) = 0$ $(x, y \in X)$, and the triangle inequality holds, i.e. $d(x,y) + d(y,z) \geqq d(x,z)$ $(x, y, z \in X)$; d is a *quasi-metric* if, in addition, $d(x,y) = 0$ implies $x = y$ $(x, y \in X)$. If d is a quasi-(pseudo)metric then so is d^{-1} defined by $d^{-1}(x,y) = d(y,x)$. A *pseudometric*

1980 *Mathematics Subject Classification* (1985 Revision). Primary 54C20, 54E35; Secondary 54E15, 54E55.

Key words and phrases. Quasi-(pseudo)metric, (weakly) bounded, compatible/continuous extension, round/stable filter, quasi-uniformity, loose extension, bitopology.

Research supported by Hungarian National Foundation for Scientific Research Grant no. 1807.

Akadémiai Kiadó, Budapest

is a symmetrical quasi-pseudometric, i.e. for which $d^{-1} = d$. In the topology induced by the quasi-pseudometric d, the sets $B_\varepsilon^d(x) = \{y: d(x,y) < \varepsilon\}$ $(\varepsilon > 0)$ form a neighbourhood base at the point $x \in X$; a *ball* is a set of the form $B_\varepsilon^d(x)$. In a topological space $X = (X, \mathcal{T})$, d is *compatible* if it induces the topology $\mathcal{T}$; it is *continuous* if it induces a topology coarser than $\mathcal{T}$ (this means continuity in the second variable at (x,x)). For $S \subset X$, let $B_\varepsilon^d(S) = \bigcup_{x \in S} B_\varepsilon^d(x)$ and $d(S,y) = \inf_{x \in S} d(x,y)$.

$d \mid S$ denotes the restriction of d to S (in fact to $S \times S$). The restriction of a continuous/compatible (quasi)-(pseudo)metric has the same property with respect to the subspace topology. A quasi-pseudometric e on $Y \supset X$ is an *extension* of d if $e \mid X = d$. If Y is a topological space then the *trace filter* of $p \in Y$ is the trace on X of the neighbourhood filter of p. (The zero filter $\exp X$ has to be allowed here.) A quasi-pseudometric d on X is *weakly bounded* if X is a ball; it follows from the triangle inequality that in this case there is for each $x \in X$ a number $n(x)$ such that $X = B_{n(x)}^d(x)$. A filter $\mathfrak{f}$ on X is *d-round* if for any $S \in \mathfrak{f}$ there are $T \in \mathfrak{f}$ and $\varepsilon > 0$ such that $B_\varepsilon^d(T) \subset S$.

Concerning quasi-uniformities, see any of [13, 8, 9]. The quasi-uniformity induced by d is denoted by $\mathcal{U}(d)$. $U_\varepsilon(d) = \{(x,y): d(x,y) < \varepsilon\}$. **N** is the set of the positive integers, **R** the set of the real numbers.

0.2 Necessary conditions. Let Y be a topological space, $\emptyset \neq X \subset Y$, d a continuous quasi-pseudometric on X. It is clearly a necessary condition for the existence of a continuous extension of d that each trace filter should contain a ball. The trace filters have to be round as well in case there is a compatible extension. Finally, a trivial necessary condition: if there exists a compatible quasi-(pseudo)metric extension then Y is, of course, quasi-(pseudo)metrizable. It will turn out that these conditions are sufficient only in some special cases.

§ 1. Extending bounded quasi-metrics

1.1 We begin with a construction in which boundedness is not assumed. Let Y be a topological space, $\emptyset \neq X \subset Y$, d a quasi-pseudometric on X, and e on Y. Generalizing a definition given in [6] for metrics and in [22] for pseudometrics, put

$$d_e(a,b) = \min\Big\{e(a,b), \inf\{e(a,x) + d(x,y) + e(y,b): x,y \in X\}\Big\} \quad (a,b \in Y).$$

LEMMA. *Assume that $d \leqq e \mid X$. Then d_e is a quasi-pseudometric on Y, $d_e \mid X = d$ and $d_e \leqq e$. If e is continuous then so is d_e. If d and e are both compatible quasi-metrics and the trace filters are d-round then d_e is a compatible quasi-metric, too.*

REMARK. It is enough to assume that the trace filters of the points in $\operatorname{cl} X \setminus X$ are round, since the trace filters belonging to the points in X (i.e. the neighbourhood filters in the subspace) are always round with respect to a compatible quasi-pseudometric, while the zero filter is evidently round. In particular, the condition on roundness is vacuous in case X is closed.

PROOF. An elementary calculation gives the triangle inequality (see in [6], where symmetry is assumed but not used). $d_e \mid X = d$, $d_e \leqq e$ and the statement on continuity are evident.

Assume now that d and e are compatible quasi-metrics and the trace filters are round. To prove the compatibility of d_e, we have to show that if $a \in Y$ and G is a neighbourhood of a then G contains a ball $B_\varepsilon^{d_e}(a)$. (It is then clear that d_e is a quasi-metric, since the existence of the compatible quasi-metric e implies that Y is a T_1-space.)

Let $\mathfrak{f}$ be the trace filter of a. Using the compatibility of e, pick $\delta > 0$ with $B_{2\delta}^e(a) \subset G$. As $\mathfrak{f}$ is round, there is a positive $\varepsilon < \delta$ such that

$$B_\varepsilon^d\left(B_\varepsilon^e(a) \cap X\right) \subset B_\delta^e(a) \cap X. \tag{1}$$

Now $B_\varepsilon^{d_e}(a) \subset B_{2\delta}^e(a) \subset G$. Indeed, if $d_e(a,b) < \varepsilon$ then either $e(a,b) < \varepsilon < 2\delta$ or there are $x, y \in X$ such that $e(a,x) + d(x,y) + e(y,b) < \varepsilon$, and in this case (1) implies $e(a,y) < \delta$, hence $e(a,b) < \delta + \varepsilon < 2\delta$. □

1.2 LEMMA. *Let d be a compatible quasi-metric and d' a quasi-pseudometric on X, $d' \geqq d$. Assume that Y is quasi-metrizable, d' has a continuous extension to Y, and the trace filters are d-round. Then d has a compatible quasi-metric extension.*

REMARK. It follows from the conditions that d' is a compatible quasi-metric, too. We shall only use this lemma in the special case when $d = d'$.

PROOF. Let d'' be a continuous extension of d', and take a compatible quasi-metric e_0 on Y. Now Lemma 1.1 can be applied with $e = e_0 + d''$. □

1.3 LEMMA. *If d is a bounded quasi-pseudometric on X and $\mathcal{U}(d)$ has a continuous extension to Y then so has d.*

PROOF. We may assume without loss of generality that $d < 1$. Let $\mathcal{V}$ be a continuous extension of $\mathcal{U}(d)$, and choose inductively $V_n \in \mathcal{V}$ $(n \geqq 0)$ such that $V_{n+1}^3 \subset V_n$ and $V_n \mid X \subset U_{2^{-n}}(d)$. The Metrization Lemma ([17] 6.12) gives a quasi-pseudometric e satisfying $V_n \subset U_{2^{-n}}(e) \subset V_{n-1}$ $(n \in \mathbf{N})$. e is clearly continuous. Now $U_{2^{-n}}(e \mid X) \subset U_{2^{-n+1}}(d)$, thus $4e \mid X \geqq d$, and Lemma 1.1 can be applied to $4e$ instead of e. □

1.4 LEMMA. *Let d be a bounded compatible quasi-metric on a subspace of a quasi-metrizable space, and assume that the trace filters are d-round. Then the following conditions are equivalent:*

(i) *d has a compatible extension;*

(ii) *d, regarded as a quasi-pseudometric, has a continuous extension;*
(iii) *$\mathcal{U}(d)$ has a compatible extension;*
(iv) *$\mathcal{U}(d)$ has a continuous extension.*

PROOF. (i)⇒(iii)⇒(iv): Evident. (iv)⇒(ii): Lemma 1.3. (ii)⇒(i): Lemma 1.2. □

This lemma implies that the results known about extensions of quasi-uniformities [7 to 11] have corollaries for *bounded* quasi-metrics. In particular, we can obtain through this lemma all the theorems of the next two sections in the special case when the quasi-metric is bounded.

§ 2. Extensions from closed or open subspaces

2.1 Closed subspace. LEMMA. *If X is a closed subspace of Y, d is a continuous quasi-pseudometric on X, and d^{-1} is weakly bounded then d has a continuous extension to Y.*

PROOF. Fix a point $x_0 \in X$, and choose $t > 0$ with $B_t^{d^{-1}}(x_0) = X$. Now

$$d'(a,b) = \begin{cases} d(a,b) & \text{if } a,b \in X, \\ 0 & \text{if } b \in Y \setminus X, \\ t + d(x_0,b) & \text{if } a \in Y \setminus X, b \in X \end{cases}$$

defines a continuous extension. □

THEOREM. *If d is a compatible quasi-metric on a closed subspace of a quasi-metrizable space, and d^{-1} is weakly bounded then d has a compatible extension.*

PROOF. Lemmas 2.1 and 1.2. □

EXAMPLES. Weak boundedness cannot be dropped from the above theorem.

a) A simple non-Hausdorff example. Let $X = \mathbf{N}$, $Y \setminus X$ the set of all the infinite sequences in $\mathbf{N}$, and take the following quasi-metric on Y: $e(n,p) = 1/p_n$ if $n \in X$, $p \in Y \setminus X$, and p_n denotes the nth element of p; otherwise $e(a,b) = 1$ for $a \neq b$. Let the topology of Y be the one induced by e. Now the metric d on X defined by $d(n,m) = |n - m|$ has no compatible quasi-metric extension.

Indeed, assume that d'' is a compatible extension, and take a point $p \in \bigcap_{n=1}^{\infty} B_1^{d''}(n)$ (there exists such a point, since $d''(n,p) < 1$ if p_n is large enough). With $n > d''(p,1) + 2$, the triangle inequality will not hold for the points n, p and 1.

b) A more complicated regular example. Let $\mathbf{Q}$ be the set of the rational numbers, $\mathbf{I} = \mathbf{R} \setminus \mathbf{Q}$, $\mathcal{E}$ the Euclidean topology on $\mathbf{R}$. Assign to each $s \in \mathbf{Q}$ a

series $G(s,n)$ of disjoint $\mathcal{E}$-open sets such that s is in the $\mathcal{E}$-closure of $G(s,n)$ ($s \in \mathbf{Q}$, $n \in \mathbf{N}$). For each pair (s,n), let $A(s,n)$ be a maximal almost disjoint collection of sequences in $G(s,n) \cap \mathbf{I}$ that $\mathcal{E}$-converge to s. Put

$$X = \bigcup\{A(s,n) : s \in \mathbf{Q}, n \in \mathbf{N}\}, \qquad Y = X \cup \mathbf{I}.$$

X is clearly almost disjoint, so the usual topology on Y is regular (in fact, zero-dimensional): let the points of $\mathbf{I}$ be isolated, and for $x \in X$, let

$$\left\{\{x\} \cup (x \setminus F) : F \text{ is finite}\right\}$$

be a neighbourhood base of x. It is easy to see that Y is quasi-metrizable. (Construct a compatible quasi-metric directly, or apply Theorem 3.1 with the role of X and $Y \setminus X$ interchanged.) X is a closed subspace, and the topology of X is discrete. Consider the following compatible quasi-metric on X:

$$d(x,y) = \max\{n-k, 1\} \quad \text{if } x \in A(s,n),\ y \in A(t,k),\ x \neq y.$$

d is weakly bounded, but it has no compatible extension.

Indeed, assume that d has a compatible extension d'', and fix points $p_0 \in Y \setminus X$ and $x_0 \in X$. It follows from the Baire category theorem that there are an $m \in \mathbf{N}$ and an interval $]u,v[$ such that $d''(p,p_0) < m$ holds for $p \in D$ where D is an $\mathcal{E}$-dense subset of $\mathbf{I} \cap]u,v[$. Pick now an $s \in \mathbf{Q} \cap]u,v[$. Choose for each $n \in \mathbf{N}$ a sequence $S_n \subset D \cap G(s,n)$ that $\mathcal{E}$-converges to s. It follows from the maximality of $A(s,n)$ that the series S_n clusters to some $y_n \in A(s,n)$, so there is a $q_n \in S_n$ with $d''(y_n,q_n) < 1$. From the triangle inequality we have

$$d(y_n,x_0) = d''(y_n,x_0) \leqq d''(y_n,q_n) + d''(q_n,p_0) + d''(p_0,x_0) < 1 + m + d''(p_0,x_0),$$

a contradiction, because $y_n \in A(s,n)$ implies that $d(y_n,x_0) \to \infty$.

2.2 Open subspace. LEMMA. *If X is an open subspace of the topological space Y, and d is a weakly bounded continuous quasi-pseudometric on X then d has a continuous extension to Y.*

PROOF. Assume that $B_t^d(x_0) = X$ and let

$$d'(a,b) = \begin{cases} d(a,b) & \text{if } a,b \in X, \\ 0 & \text{if } a \in Y \setminus X, \\ t + d(a,x_0) & \text{if } a \in X,\ b \in Y \setminus X. \end{cases} \qquad \square$$

THEOREM. *If d is a weakly bounded compatible quasi-metric on an open subspace of a quasi-metrizable space, and the trace filters are d-round then d has a compatible extension.*

PROOF. Lemmas 2.2 and 1.2. $\square$

PROBLEM. Assume that d is a compatible quasi-metric on a (dense) open subspace of a quasi-metrizable space, each trace filter is d-round and contains a ball. Does d have a compatible extension?

§ 3. Extensions from dense subspaces

3.1 Loose extensions. Let us now consider the case when Y is a *loose extension* of X belonging to trace filters $\mathfrak{f}(a)$ $(a \in Y)$, which means that $\{\{a\} \cup S : S \in \mathfrak{f}(a)\}$ is a neighbourhood base of a in Y. X is open in Y, so 2.2 can be applied; more is true, however: we are going to show that the answer to Problem 2.2 is positive in this special case.

LEMMA. *Assume that Y is a loose extension of X, d is a continuous quasi-pseudometric on X, and each trace filter contains a ball. Then d has a continuous extension to Y.*

PROOF. Fix balls $S_p \in \mathfrak{f}(p)$ $(p \in Y \setminus X)$; for $x \in X$, let $S_x = \{x\}$. Define now

$$d'(a,b) = \inf\{t \in \mathbf{R} : S_b \subset B_t(S_a)\}. \qquad \square$$

THEOREM. *If d is a compatible quasi-metric on X, Y is a loose extension of X and it is a first countable T_1-space, the trace filters are round, and each trace filter contains a ball then d has a compatible extension to Y.*

PROOF. It is enough to show that Y is quasi-metrizable, because then Lemmas 3.1 and 1.2 can be applied.

$\mathcal{U} = \mathcal{U}(d)$ has a countable base, the trace filters are $\mathcal{U}$-round (because $\mathcal{U}$-round = d-round), and Y is first countable; according to [10] 6.5, these conditions are sufficient for the existence of a compatible extension $\mathcal{V}$ having a countable base. $\mathcal{V}$ is quasi-pseudometrizable (e.g. [18] 11.1.1), so T_1 implies that $\mathcal{V}$ is quasi-metrizable. $\square$

The example below shows that the quasi-metrizability of Y cannot be replaced in Theorem 2.2 by the assumption that Y is a first countable T_1-space (not even when X is dense and d is bounded).

EXAMPLE. Let Z_1 be a first countable T_1-space that is not quasi-metrizable (see [13] for such spaces), and $Z_2 = \mathbf{N} \cup \{\omega\}$ a convergent sequence with the limit point ω. Take $Y = Z_1 \times Z_2$, $X = Z_1 \times \mathbf{N}$, and modify the product topology on Y by declaring the points of X isolated. Consider the (quasi-)metric d on X defined by $d(x,y) = 1$ if $x \neq y$. The trace filters are clearly round, d is compatible, X is dense and open, Y is first countable and T_1, but it is not quasi-metrizable, because its subspace $Y \setminus X$ is homeomorphic with Z_1.

3.2 Other extensions. Recall that a filter $\mathfrak{f}$ is *d-stable* (where d is a quasi-pseudometric) [8, 10] if for any $\varepsilon > 0$, $\bigcap_{S \in \mathfrak{f}} B_\varepsilon(S) \in \mathfrak{f}$.

LEMMA. *If X is a dense subspace of the topological space Y, d is a continuous quasi-pseudometric on X, the trace filters are stable and each trace filter contains a ball then d has a continuous extension to Y.*

PROOF. Define

$$d'(a,b) = \inf\{t \in \mathbf{R} : S \in \mathfrak{f}(a) \Rightarrow B_t^d(S) \in \mathfrak{f}(b)\}.$$

$d'(a,b)$ is finite: pick a point $x \in \bigcap_{S \in \mathfrak{f}(a)} B_1^d(S)$; now $B_s^d(x) \in \mathfrak{f}(b)$ for a suitable $s \in \mathbf{R}$ (because $\mathfrak{f}(b)$ contains a ball), thus $d'(a,b) < s+1$. d' is clearly a quasi-pseudometric with $d' \mid X = d$. To prove that d' is continuous, fix $a \in Y$ and $\varepsilon > 0$; we have to show that $B_\varepsilon^{d'}(a)$ is a neighbourhood of a in the original topology of Y. Using the stability of $\mathfrak{f}(a)$, take an open neighbourhood G of a such that $G \cap X \subset \bigcap_{S \in \mathfrak{f}(a)} B_{\varepsilon/2}^d(S)$. For each $b \in G$, $G \cap X \in \mathfrak{f}(b)$, thus $G \subset B_\varepsilon^{d'}(a)$. □

THEOREM. *Let d be a compatible quasi-metric on a dense subspace of a quasi-metrizable space, assume that the trace filters are d-round and d-stable, and each trace filter contains a ball. Then d has a compatible extension.*

PROOF. Lemmas 3.2 and 1.2. □

REMARK. If Y is a strict extension of X (i.e. the coarsest one belonging to the given trace filters) then the quasi-metrizability of Y can be replaced in the above theorem by the assumption that Y is a T_1-space. In this case the quasi-pseudometric d' defined in the proof of the lemma is already a compatible quasi-metric on Y if d satisfies the conditions of the theorem.

§ 4. The bitopological case

A quasi-pseudometric d is *compatible/continuous* in the bitopological space $X = (X; \mathcal{T}^{-1}, \mathcal{T}^1)$ if d^i is compatible/continuous in $(X, \mathcal{T}^i)$ $(i = \pm 1)$, where $d^1 = d$. Now if X is a subspace of the bitopological space Y then we may ask whether d has a continuous or compatible extension to Y. There is one more necessary condition besides the ones in 0.2: if a point $p \in Y \setminus X$ belongs to the closure of X in both topologies then the trace filter pair $(\mathfrak{f}^{-1}(p), \mathfrak{f}^1(p))$ has to be d-Cauchy. (A filter pair $(\mathfrak{f}^{-1}, \mathfrak{f}^1)$ is *d-Cauchy* if for any $\varepsilon > 0$ there are sets $S_i \in \mathfrak{f}^i$ such that $d(x_{-1}, x_1) < \varepsilon$ whenever $x_i \in \mathfrak{f}^i$ $(i = \pm 1)$.)

Sufficient conditions can be obtained (at least when d is bounded) from the results known about extending quasi-uniformities in bitopological spaces, see [9, 10] and [11] § 2. For this purpose, observe that the bitopological analogue of Lemma 1.4 is valid (where Lemma 1.1 is needed in the proof, apply it both for (d, e) and (d^{-1}, e^{-1}), and check that $(d_e)^{-1} = (d^{-1})_{e^{-1}}$.

[10] Theorem 9.1 is the only result known to the author about extensions of unbounded quasi-pseudometrics in bitopological spaces.

ADDED IN PROOF. See [24] 4.3 and [25] for more about the bitopological problem.

REFERENCES

[1] ALÒ, R. A., Results related to P-embedding, *Topics in topology* (Proc. Third Colloq., Keszthely, 1972), Colloq. Math. Soc. J. Bolyai, **8**, North-Holland, Amsterdam, 1974, 29–40. *MR* **51** #6713

[2] ALÒ, R. A., IMLER, L. and SHAPIRO, H. L., P- and z-embedded subspaces, *Math. Ann.* **188** (1970), No. 1, 13–22. *MR* **42** #1062

[3] ALÒ, R. A. and SHAPIRO, H. L., Continuous uniformities, *Math. Ann.* **185** (1970), No. 4, 322–328. *MR* **41** #4484

[4] ARENS, R., Extensions of coverings, of pseudo-metrics, and of linear-space-valued mappings, *Canadian J. Math.* **5** (1953), 211–215. *MR* **14**–1108

[5] BACON, P., Extending a complete metric, *Amer. Math. Monthly* **75** (1968), No. 6, 642–643. *MR* **37** #5848

[6] BING, R. H., Extending a metric, *Duke Math. J.* **14** (1947), 511-519. *MR* **9**–521

[7] CSÁSZÁR, Á., Regular extensions of quasi-uniformities, *Studia Sci. Math. Hungar.* **14** (1979), No. 1–3, 15–26. *MR* **83i:** 54026

[8] CSÁSZÁR, Á., Extensions of quasi-uniformities, *Acta Math. Acad. Sci. Hungar.* **37** (1981), No. 1–3, 121–145. *MR* **82f:** 54039

[9] DEÁK, J., Extensions of quasi-uniformities for prescribed bitopologies I, *Studia Sci. Math. Hungar.* **25** (1990), No. 1–2, 45–67. *MR* **92b**: 54058

[10] DEÁK, J., Extensions of quasi-uniformities for prescribed bitopologies II, *Studia Sci. Math. Hungar.* **25** (1990), No. 1–2, 69–91. *MR* **92b**: 54058

[11] DEÁK, J., Quasi-uniform extensions for finer topologies, *Studia Sci. Math. Hungar.* **25** (1990), No. 1–2, 97–105. *MR* **92b**: 54059

[12] DOWKER, C. H., On a theorem of Hanner, *Ark. Mat.* **2** (1952), No. 4, 307–313. *MR* **14**–396

[13] FLETCHER, P. and LINDGREN, W. F., *Quasi-uniform spaces*, Lecture Notes in Pure Appl. Math., **77**, Marcel Dekker, New York, 1982. *MR* **84h:** 54026

[14] GANTNER, T. E., Extensions of uniformly continuous pseudometrics, *Trans. Amer. Math. Soc.* **132** (1968), No. 1, 147–157. *MR* **36** #5886

[15] GANTNER, T. E., Extensions of uniform structures, *Fund. Math.* **66** (1969/1970), No. 3, 263–281. *MR* **42** #5220

[16] HAUSDORFF, F., Erweiterung einer Homöomorphie, *Fund. Math.* **16** (1930), 353–360. *Jb. Fortschritte Math.* **56**, 508

[17] KELLEY, J. L., *General topology*, D. Van Nostrand Company, Inc., Toronto–New York–London, 1955. *MR* **16**–1136

[18] PERVIN, W. J., *Foundations of general topology*, Academic Press, New York, 1964. *MR* **29** #2759

[19] SHAPIRO, H. L., Extensions of pseudometrics, *Canad. J. Math.* **18** (1966), No. 5, 981–998. *MR* **34** #6719

[20] SHAPIRO, H. L., More on extending continuous pseudometrics, *Canad. J. Math.* **22** (1970), No. 5, 984–993. *MR* **43** #2665

[21] TORUŃCZYK, H., A short proof of Hausdorff's theorem on extending metrics, *Fund. Math.* **77** (1972), No. 2, 191–193. *MR* **47** #9559

[22] ÚRY, L., Extending compatible uniformities, *Topology* (Proc. Fourth. Colloq., Budapest, 1978), Vol. II, Colloq. Math. Soc. J. Bolyai **23**, North-Holland, Amsterdam–New York, 1980, 1185–1209. *MR* **82g:** 54043

[23] WILSON, W. A., On quasi-metric spaces, *Amer. J. Math.* **53** (1931), 675–684. *Zbl* **2**–55

[24] DEÁK, J., A survey of compatible extensions (presenting 77 unsolved problems), *Topology, theory and applications II* (Proc. Sixth Colloq., Pécs, 1989), Colloq. Math. Soc. J. Bolyai **55**, North-Holland, Amsterdam, 1993, 127–175.

[25] SALBANY, S. and ROMAGUERA, S., The Hausdorff extension theorem for distance spaces (manuscript, 1992).

(Received June 10, 1989)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 115–129

ON AN ALGEBRAIC DIFFERENTIAL EQUATION OF BERNOULLI TYPE

T. FÉNYES

Introduction

In paper [1] we discussed the algebraic Bernoulli type differential equation

$$D(x) + ax + bx^m = 0 \tag{1}$$

in the discrete Mikusiński operator field based on the Cauchy product of functions. Here m is an arbitrary integer ($m \neq 0, 1$), a, b are arbitrarily given real valued functions defined on the set of the non-negative integers, D is the symbol of the algebraic derivative.

In the present paper we shall deal with (1) in the case where the operator field is based on the number-theoretical Dirichlet product of functions defined on the set of the positive integers.

The paper consists of three chapters. In Chapter 1 we summarize the elements and known results of the operational calculus based on the Dirichlet product (see [2], [3], [4]). Chapter 2 contains the operational theory of (1) in the case of $m = 2$.

The more complicated cases for $m \neq 2$ are treated in Chapter 3. In what follows Z, R will denote the sets of natural and positive rational numbers, respectively.

§1. Discrete Mikusiński operators based on the Dirichlet product

Let $a = \{a(n)\}$ be an arbitrary real-valued function defined on Z. The symbol $a(n)$ denotes the value of this function for arbitrary fixed n.

Let E denote the set of the discrete functions. If we introduce in E the following two operations

(i) $a + b$: $\{a(n)\} + \{b(n)\} = \{a(n) + b(n)\}$ addition,

(ii) ab: $\{a(n)\}\{b(n)\} = \left\{\sum_{\nu|n} a(\nu)b(\frac{n}{\nu})\right\}$ multiplication,

1991 *Mathematics Subject Classification* (1985 Revision). Primary 44A40; Secondary 11A99, 13N99.

Key words and phrases. Operational calculus, number theory.

Research partially supported by Hungarian National Foundation for Scientific Research Grant no. 6032/6319.

then E becomes a commutative ring without divisor of zero and can be extended to a quotient field. This is called the discrete Mikusiński operator field and is denoted by M_D. The elements of M_D are called M_D-operators.

The definition and properties of the "discrete" Dirac-function

We define the discrete Dirac-function by

$$\delta(N) = \{\delta(n, N)\},$$

where

$$\delta(n, N) = \begin{cases} 0, & \text{for } n \neq N, \\ 1, & \text{for } n = N. \end{cases}$$

For later purposes we enumerate some properties of the Dirac function.

PROPERTY 1. $$\delta(N)\{a(n)\} = \{b(n)\},$$

$$b(n) = \begin{cases} a\left(\frac{n}{N}\right), & \text{for } N \mid n, \\ 0 & \text{otherwise.} \end{cases} \tag{1.1}$$

$$\delta(N_1)\delta(N_2) = \delta(N_1 N_2), \quad N_1, N_2 \in Z. \tag{1.2}$$

PROPERTY 2.

$$x = \frac{\{a(n)\}}{\delta(N)} \in E, \qquad N \in Z \tag{1.3}$$

holds if and only if

$$a(n) = 0 \tag{1.4}$$

for those values of n for which N is not a divisor of n. If (1.4) holds, then

$$x = \{a(nN)\}. \tag{1.5}$$

The field K of the real or complex numbers can be embedded isomorphically into the operator field M_D. The common unit element of K, E, M_D is the function $\delta(1)$ and we write

$$\delta(1) = 1.$$

Moreover,

$$c\delta(1) = c, \quad c\{a(n)\} = \{ca(n)\};$$
$$\text{for every } c \in K \text{ and every } a \in E.$$

Every operator of the form

$$x = \frac{\{a(n)\}}{\{b(n)\}}$$

is a function if $b(1) \neq 0$. Moreover $\frac{1}{\{b(n)\}} \in E$ iff $b(1) \neq 0$.

The operator function $\delta(\varepsilon)$

For arbitrary rational number $\varepsilon = \frac{N_1}{N_2} \in R$ we define

$$\delta(\varepsilon) = \frac{\delta(N_1)}{\delta(N_2)}. \tag{1.6}$$

From this definition it follows that for $\varepsilon = N$ we have

$$\delta(\varepsilon) = \delta(N) = \{\delta(n, N)\}.$$

If

$$\frac{N_1}{N_2} = \frac{N_3}{N_4},$$

then

$$\delta\left(\frac{N_1}{N_2}\right) = \delta\left(\frac{N_3}{N_4}\right)$$

holds.

PROPERTY 3. Let α, β be arbitrary positive rational numbers, then

$$\delta(\alpha)\delta(\beta) = \delta(\alpha\beta) \tag{1.7}$$

and it is easily seen that

$$\delta\left(\frac{1}{\alpha}\right) = \frac{1}{\delta(\alpha)} \tag{1.8}$$

is also true.

The definition of the ring E^*

Let $E^* \subset M_D$ be the subset of M_D whose elements are of the form

$$x = \frac{a}{\delta(N)} \qquad N \in Z,\ a \in E. \tag{1.9}$$

E^* is a ring-and, by choosing $N = 1$, we have

$$E \subset E^*.$$

PROPERTY 4. Obviously,

$$x = \frac{a}{\delta(\varepsilon)} \in E^*, \qquad \varepsilon = \frac{N_1}{N_2} \quad (N_1, N_2 \text{ are relatively primes}).$$

Moreover, $x \in E$ if and only if

$$a(n) = 0$$

for those values of n for which N_1 is not a divisor of n. If the condition is satisfied, we have

$$x(n) = \begin{cases} a\left(\frac{nN_1}{N_2}\right), & \text{for } N_2 \mid n, \\ 0 & \text{otherwise.} \end{cases}$$

Definition of the convergence in the ring E

Let $a_k \in E$, $(k = 1, 2, \dots)$ be an infinite sequence of functions. By definition

$$\lim_{k\to\infty} \{a_k(n)\} = \{a(n)\} \tag{1.10}$$

if for every fixed n

$$\lim_{k\to\infty} a_k(n) = a(n).$$

This convergence can be extended to infinite series of functions as usual.

Let

$$f(z) = \sum_{k=0}^{\infty} \beta_k z^k, \quad \beta_k \in K$$

be an arbitrary entire function of the complex variable z. Then

$$f(a) = \sum_{k=0}^{\infty} \beta_k \{a(n)\}^k, \quad a \in E,\ a^0 = 1 \tag{1.11}$$

holds in the sense of the convergence defined above. We have

$$e^a = \sum_{k=0}^{\infty} \frac{a^k}{k!}, \quad a \in E,\ a^0 = 1$$

having the property

$$e^a e^b = e^{a+b}, \qquad a, b \in E, \tag{1.11'}$$

moreover, if we write

$$e^a = \{e_a(n)\}$$

so

$$e_a(1) = e^{a(1)} \tag{1.12}$$

holds. Moreover, let

$$\sum_{k=0}^{\infty} \gamma_k z^k, \qquad \gamma_k \in K \tag{1.12'}$$

be an arbitrary *formal* infinite series and let $a \in E$ an arbitrary function with $a(1) = 0$. Then

$$\sum_{k=0}^{\infty} \gamma_k a^k \tag{1.12''}$$

also converges in the sense of convergence defined above.

The algebraic derivation and integration

For the sake of easy reading we recapitulate some definitions and facts of the algebraic derivation and integration.

$$D(a) = \{-\log n\, a(n)\}, \qquad a \in E \tag{1.13}$$

$$D\left(\frac{a}{b}\right) = \frac{bD(a) - aD(b)}{b^2}, \quad a, b \in E, \ \frac{a}{b} \in M_D. \tag{1.13'}$$

PROPERTY 5.

$$D\left[\frac{a}{\delta(\varepsilon)}\right] = \frac{\{-\log \frac{n}{\varepsilon}\, a(n)\}}{\delta(\varepsilon)} \in E^*, \ a \in E, \ \varepsilon = \frac{N_1}{N_2}. \tag{1.14}$$

$$D[\delta(\varepsilon)] = -\log \varepsilon\, \delta(\varepsilon). \tag{1.14'}$$

PROPERTY 6.

$$D(e^a) = D(a)e^a, \qquad a \in E. \tag{1.15}$$

If for a given $x \in M_D$ there exists a $y \in M_D$ such that

$$D(y) = x$$

we say that x is algebraic integrable and we write

$$y = \int x.$$

PROPERTY 7. If $x \in M_D$ and

$$D(x) = 0$$

then x is an arbitrary complex number.

Two algebraic integrals of an operator may differ only by an arbitrary number.

The algebraic differentiation and integration is a linear operation over the field of the real (complex) numbers.

PROPERTY 8. The operator

$$x = \frac{a}{\delta(\varepsilon)}, \quad a \in E, \quad \varepsilon = \frac{N_1}{N_2} \in R \tag{1.16}$$

is algebraic integrable in M_D if and only if either $\varepsilon \neq N$, $N \in Z$, or $\varepsilon = N$ and $a(N) = 0$ holds true. Every algebraic integral of (1.16) belonging to E^* is given by

$$\int \frac{a}{\delta(\varepsilon)} = \frac{\left\{-\frac{a(n)}{\log \frac{n}{\varepsilon}}\right\}}{\delta(\varepsilon)} + c, \quad c \in K \tag{1.17}$$

where in the case of $\varepsilon = N$ the symbol $\frac{a(N)}{\log \frac{N}{N}}$ denotes an arbitrary real (complex) number. We shall choose this to be null.

For $\varepsilon = 1$ we have that a is integrable if and only if $a(1) = 0$, and

$$\int a = \left\{-\frac{a(n)}{\log n}\right\}.$$

Let us consider the differential equation

$$D(x) - fx = h \qquad f, h \in E \tag{1.18}$$

with respect to which the following theorem holds.

THEOREM (see [2], [3]). *The homogeneous equation*

$$D(x) - fx = 0 \tag{1.19}$$

has a nontrivial solution in M_D if and only if

$$\alpha = e^{-f(1)} \in R.$$

The general solution of (1.19) *is of the form*

$$x = c\delta(\alpha)\exp\left[\int (f - f(1))\right], \quad c \in K \tag{1.20}$$

being a function for $c \neq 0$ if and only if $\alpha \in Z$. (1.18) *has a solution $x_p \in M_D$ if and only if one of the following conditions holds.*

(1.20′) (i) $\alpha = e^{-f(1)} \notin Z$ *or* (ii) $\alpha = e^{-f(1)} \in Z, \quad H(\alpha) = 0,$

where

$$H = \{H(n)\} = he^{-\int (f - f(1))}.$$

Moreover

$$x_p = \left\{\frac{-H(n)}{\log n + f(1)}\right\}\exp\left[\int (f - f(1))\right] \in E, \tag{1.21}$$

where in the case of (ii) *the symbol*

$$\frac{H(\alpha)}{\log \alpha + f(1)} = \frac{0}{0} \tag{1.22}$$

denotes the number zero.

§2. The case $m = 2$

Let us consider the Bernoulli equation

$$D(x) + ax + bx^2 = 0, \qquad a, b \in E, \quad b \neq 0. \tag{2.1}$$

By the application of the substitution

$$x = \frac{1}{z}$$

(2.1) can be reduced to the linear equation of the form

$$D(z) - az = b. \tag{2.2}$$

We extend the definition of $\delta(\alpha)$ for irrational α by $\delta(\alpha) = 0$. So we have the following

THEOREM 1. *If $e^{-a(1)} \notin Z$, then the general solution of* (2.1) *is of the form*

$$x = \left[c\delta(e^{-a(1)}) - \left\{\frac{G(n)}{\log n + a(1)}\right\}\right]^{-1} e^{-\int(a-a(1))}, \quad c \in K \tag{2.3}$$

where

$$G = be^{-\int(a-a(1))}. \tag{2.4}$$

If $e^{-a(1)} \in Z$, then (2.1) *has a nontrivial solution in M_D if and only if $G(e^{-a(1)}) = 0$. If so, then* (2.3) *is the general solution of* (2.1), *where*

$$\frac{G(e^{-a(1)})}{\log e^{-a(1)} + a(1)} = 0.$$

Moreover, if $e^{-a(1)} \notin R$, then $x \in E$, iff $b(1) \neq 0$. If $e^{-a(1)} = \frac{1}{N}$, $N \in Z$, $(N \geqq 2)$, then $x \in E$ for every $c \neq 0$ and for $c = 0$, $x \in E$ iff $b(1) \neq 0$.

If $e^{-a(1)} = \frac{M}{N}$, (M and N are relatively primes), $M > 1$, $N > 1$, then $x \in E$ iff $c = 0$ and $b(1) \neq 0$. Finally, let $e^{-a(1)} = N$, $N \in Z$ and $G(N) = 0$. If $N = 1$, then $x \in E$ iff $c \neq 0$. If $N > 1$, then $x \in E$ iff $b(1) \neq 0$.

PROOF. Taking into account the Theorem of the preceding chapter, we can see that only the existence criteria of nontrivial solutions $x \in E$ of (2.1) are to be proved.

Since

$$e^{f_1} e^{f_2} = e^{f_1 + f_2}, \qquad f_1, f_2 \in E,$$

it is obvious that $x \in E$ iff

$$y = \left(c\delta(e^{-a(1)}) - \left\{ \frac{G(n)}{\log n + a(1)} \right\} \right)^{-1} \in E. \tag{2.5}$$

I. If $e^{-a(1)} \notin R$, then $\delta(e^{-a(1)}) = 0$. Since by (1.12), (2.4)

$$G(1) = b(1)$$

it is easily seen by Property 2 that $y \in E$ iff $b(1) \neq 0$.

II. If $e^{-a(1)} = \frac{1}{N}$, $(N \geqq 2)$, we write

$$y = \frac{\delta(N)}{c - \delta(N) \left\{ \frac{G(n)}{\log n + a(1)} \right\}}. \tag{2.6}$$

From Property 1 and 2 it follows that for $c \neq 0$ $x \in E$ holds. For $c = 0$ the case I is obtained.

III. If $e^{-a(1)} = \frac{M}{N} \notin Z$, $M > 1$ we write

$$y = \frac{\delta(N)}{c\delta(M) - \delta(N) \left\{ \frac{G(n)}{\log n + a(1)} \right\}}. \tag{2.7}$$

For $c = 0$ the case I is obtained again. Let $c \neq 0$. We show that $y \notin E$. Let us suppose the contrary. If $y \in E$, then by Property 1

$$cy\left(\frac{n}{M}\right) - g\left(\frac{n}{N}\right) = \delta(N), \qquad n = 1, 2, \ldots \tag{2.8}$$

holds, where

$$g = y\left\{ \frac{G(n)}{\log n + a(1)} \right\}.$$

For $n = N$ we have

$$-g(1) = -y(1)\frac{b(1)}{a(1)} = 1,$$

and for $n = M$

$$cy(1) = 0.$$

Consequently $y(1) = 0$ and $0 = 1$, a contradiction.

IV. Let $e^{-a(1)} = N$, $N \in Z$. If $N > 1$, then the value of the function

$$c\delta(N) - \left\{ \frac{G(n)}{\log n + a(1)} \right\}$$

equals to

$$-\frac{G(1)}{a(1)} = -\frac{b(1)}{a(1)}$$

for $n = 1$, so by Property 2 we have that $y \in E$ for every c iff $b(1) \neq 0$.

If $N = 1$, then by the Theorem of the preceding chapter we have that (2.3) exists iff $b(1) = 0$. But if $b(1) = 0$, then $y \in E$ iff $c \neq 0$.

§3. The case $m \neq 2$

In the sequel we shall deal with (1) if $m \neq 2$. For $m > 2$ (1) has the trivial solution $x = 0$. If m is odd and $x_0 \in M_D$ is a solution of (1), then $-x_0$ is also a solution of (1). In the sequel we shall not distinguish between solutions differing from each other only in their signs. Moreover, we find only real solutions of (1). An $x \in M_D$ is called to be real if $x = \frac{a}{b}$, $a, b \in E$, and the functions a, b are real-valued.

By the application of the substitution

$$z = x^{1-m} \tag{3.1}$$

we can reduce (1) to the linear equation of the form

$$D(z) - (m-1)az = (m-1)b. \tag{3.2}$$

If (1) has a nontrivial solution in M_D (3.1) shows that (3.2) is also solvable in M_D. The converse statement does not hold. If (3.2) has a solution $z \in M_D$, we obtain

$$x = z^{\frac{1}{1-m}} \tag{3.3}$$

as a formal solution of (1). However, (3.3) does not exist necessarily, since the field M_D is not algebraically closed. We show this.

We define the subsets $\tilde{E}_k \subset E$ as follows ($k = 1, 2, \dots$).

DEFINITION. $z \in \tilde{E}_k$ iff

$$\frac{z}{\delta(k)} \in E$$

and $z(k) \neq 0$. By Property 2 we see that for $z \in \tilde{E}_k$

$$\frac{z}{\delta(k)} = \{z(kn)\}.$$

If $z \in E$, $z(1) \neq 0$, then $z \in \tilde{E}_1$.

LEMMA 1. *Let $q \in Z$, $(q \geqq 2)$, $p \in R$, $z \in \tilde{E}_k$. Then $\sqrt[q]{\delta(p)z} \in M_D$ exists if and only if*

$$\sqrt[q]{pk} \in R$$

and for even q, $z(k) > 0$. Moreover,

$$\sqrt[q]{\delta(p)z} = \delta\left(\sqrt[q]{pk}\right) \sqrt[q]{z(k)} \exp \frac{1}{q} \int \frac{D(u)}{u} \tag{3.2'}$$

where $u = \{z(kn)\}$.

PROOF. Let us show first that for every $f \in E$, $f(1) \neq 0$

$$f = f(1) \exp \int \frac{D(f)}{f}$$

holds. By the theorem of the first chapter it is easily seen that the differential equation

$$(*) \qquad D(w) - \frac{D(f)}{f} w = 0$$

has the general solution of the form

$$w = c \exp \int \frac{D(f)}{f}, \qquad c \in K.$$

Moreover, f is also a solution of $(*)$. Then $\exists \tilde{c} \in K$ such that

$$f = \tilde{c} \exp \int \frac{D(f)}{f}.$$

Substituting $n = 1$, we have

$$f(1) = \tilde{c}$$

so

$$f = f(1) \exp \int \frac{D(f)}{f}.$$

We obtain

$$\delta(p)z = \delta(p)\frac{z}{\delta(k)}\delta(k) = \delta(pk)u = \delta(pk)z(k)e^{\int \frac{D(u)}{u}}.$$

Let $\sqrt[q]{pk} \in R$. By $(1.11')$ it can be seen that $(3.2')$ holds true if for even q, $z(k) > 0$. Let us assume that there exists a $\tau \in M_D$ such that

$$\sqrt[q]{z\delta(p)} = \tau.$$

Then

$$\tau^q = z\delta(p) = u\delta(kp).$$

We have

$$D(\tau^q) = q\tau^{q-1}D(\tau) = -\log kp\, \delta(kp)u + \delta(kp)D(u).$$

Multiplying by τ

$$q\tau^q D(\tau) = (-\log kp\, \delta(kp)u + \delta(kp)D(u))\tau$$

and the following differential equation is obtained:

$$(3.2'') \qquad D(\tau) - \varrho\tau = 0,$$

where

$$\varrho = \frac{1}{q}\frac{D(u)}{u} - \frac{\log kp}{q}.$$

Since $u(1) = z(k) \neq 0$, so is $\frac{D(u)}{u} \in E$, $\varrho \in E$ and applying the theorem of Chapter 1

$$e^{-\varrho(1)} = e^{\frac{\log kp}{q}} = \sqrt[q]{kp} \in R$$

holds and the lemma has been proved. On the base of what has been told we pronounce

LEMMA 2. *Let $m \neq 2$. (1) has formal solutions iff*

$$e^{-(m-1)a(1)} \notin Z, \text{ or } e^{-(m-1)a(1)} \in Z \text{ and}$$

$$G_m\left(e^{-(m-1)a(1)}\right) = 0, \tag{3.2'''}$$

where

$$G_m = (m-1)be^{-\int (m-1)(a-a(1))}. \tag{3.3}$$

The general formal solution of (1) is of the form

(3.4)

$$x = \frac{\exp\left[-\int(a-a(1))\right]}{\sqrt[m-1]{c\delta\left(e^{-(m-1)a(1)}\right) - \left\{\frac{G_m(n)}{\log n + (m-1)a(1)}\right\}}}, \text{ for } m > 2$$

$$x = \sqrt[|m|+1]{c\delta\left(e^{-(m-1)a(1)}\right) - \left\{\frac{G_m(n)}{\log n + (m-1)a(1)}\int\right\}e^{-\int(a-a(1))}}, \text{ for } m < 0$$

where in the case of $e^{-(m-1)a(1)} \in Z$

$$\frac{G_m\left(e^{-(m-1)a(1)}\right)}{\log e^{-(m-1)a(1)} + (m-1)a(1)}$$

denotes the number zero.

It arises now the following question. Under what conditions do the formal solutions represent proper solutions of the Bernoulli equation?

Obviously, only in the case when the roots occurring in (3.4) exist. In the sequel we give only simple sufficient criteria guaranteeing the existence of (3.4) in M_D or E, respectively.

We need the following trivial

STATEMENT. 1. If for $T_1, T_2 \in M_D$, $q \in Z$

$$\sqrt[q]{T_1} = R_1, \quad \sqrt[q]{T_2} = R_2,$$

then $\sqrt[q]{T_1 T_2} = R_1 R_2$, $\sqrt[q]{\frac{T_1}{T_2}} = \frac{R_1}{R_2}$, especially $\sqrt[q]{\frac{1}{T_2}} = \frac{1}{\sqrt[q]{T_2}}$.

If for $Q_1, Q_2 \in M_D$, $\sqrt[g]{Q_1}$ exists and $\sqrt[g]{Q_2}$ does not, then $\sqrt[g]{Q_1 Q_2}$, $\sqrt[g]{\frac{Q_1}{Q_2}}$ do not exist in M_D.

2. Let $p_1 \in \tilde{E}_k$, $p_2 \in E$, $p_2(1) \neq 0$. Then

$$\varphi = \Big\{ \sum_{\nu|n} p_1(\nu) p_2\Big(\frac{n}{\nu}\Big) \Big\} \in \tilde{E}_k$$

and

$$\varphi(k) = p_1(k) p_2(1)$$

holds.

From Lemma 2 and the above statement we obtain the following

COROLLARY. *Let* $e^{-(m-1)a(1)} = N \in Z$ *and* $b \in \tilde{E}_k$. *If* $N = k$, *then* (1) *has no formal solution. If* k *is not a divisor of* N, *then* (3.4) *is the general formal solution of* (1).

Now we can prove the following

THEOREM 2. I. *Let* $e^{-(m-1)a(1)}$ *be irrational or* $c = 0$ *and* $b \in \tilde{E}_k$. *Let us assume that for any integer value of* $e^{-(m-1)a(1)}$ *the condition* (3.2‴), (3.3) *is satisfied. For* $m > 2$ (3.4) *exists in* M_D *if and only if* $\sqrt[m-1]{k} \in Z$ *and for odd* m

$$\frac{b(k)}{\log k + (m-1)a(1)} < 0.$$

If so, then

$$(*) \qquad \begin{aligned} x &\notin E, \quad \text{if} \quad k > 1 \\ x &\in E, \quad \text{if} \quad k = 1. \end{aligned}$$

For $m < 0$ (3.4) *exists in* M_D *if and only if* $\sqrt[|m|+1]{k} \in Z$ *and for odd* $|m|$

$$\frac{b(k)}{\log k + (m-1)a(1)} > 0.$$

If so, then

$$(**) \qquad x \in E.$$

II. *Let* $c \neq 0$ *and* $e^{-(m-1)a(1)} = \frac{1}{N}$, $N \in Z$. *For* $m > 2$ (3.4) *exists in* M_D *if and only if* $\sqrt[m-1]{N} \in Z$ *and for odd* $m, c > 0$. *If so, then*

$$x \in E.$$

For $m < 0$ (3.4) *exists in* M_D *if and only if* $\sqrt[|m|+1]{N} \in Z$ *and for odd* $|m|$, $c > 0$. *If so, then*

$$x \in E \quad \text{if} \quad N = 1.$$

$$x \notin E \quad \text{if} \quad N > 1.$$

III. *Let* $c \neq 0$ *and* $e^{-(m-1)a(1)} = N$, $(N > 1)$, $b \in \tilde{E}_k$, $(k \neq N)$. *If* $k \mid N$, *and the condition* (3.2‴), (3.3) *is satisfied, then the existence criteria of the case I and* (∗); (∗∗) *hold. If* $N \mid k$, *then the existence criteria of case II hold. Moreover, for* $m > 2$ $x \notin E$, *for* $m < 0$ $x \in E$ *holds.*

PROOF. I. Let us introduce the notation

$$\psi_m = \left\{ \frac{G_m(n)}{\log n + (m-1)a(1)} \right\}.$$

Since (3.2‴), (3.3) is satisfied, by (3.4) we have

$$x = \frac{e^{-\int (a - a(1))}}{\sqrt[m-1]{-\psi_m}}, \quad \text{for } m > 2, \tag{3.5}$$

$$x = \sqrt[|m|+1]{-\psi_m}\, e^{-\int (a-a(1))}, \quad \text{for } m < 0. \tag{3.6}$$

Since $b \in \tilde{E}_k$, it follows from the second part of the statement that $\psi_m \in \tilde{E}_k$. By (1.12), (3.3)

$$\psi_m(k) = \frac{(m-1)b(k)}{\log k + (m-1)a(1)}.$$

Applying Lemma 1, (3.2′) for $p = 1$, and the first part of the statement, we obtain that for $m > 2$ x exists in M_D if and only if

$$\sqrt[m-1]{k} \in Z \tag{3.7}$$

and for odd m

$$\frac{b(k)}{\log k + (m-1)a(1)} < 0. \tag{3.8}$$

By taking again into account (3.2′) we see that if (3.7), (3.8) hold, then $x \in E$ iff $k = 1$.

The case $m < 0$ can be proved similarly.

II. By (3.4) we have

$$x = \frac{e^{-\int (a - a(1))}}{\sqrt[m-1]{c\delta\left(\frac{1}{N}\right) - \psi_m}}, \quad m > 2, \tag{3.9}$$

$$x = \sqrt[|m|+1]{c\delta\left(\frac{1}{N}\right) - \psi_m}\, e^{-\int (a-a(1))}, \quad m < 0. \tag{3.10}$$

For arbitrary $q \in Z$

$$\sqrt[q]{c\delta\left(\frac{1}{N}\right) - \psi_m} = \sqrt[q]{\frac{c - \delta(N)\psi_m}{\delta(N)}}. \tag{3.11}$$

We have by Property 1 that $\delta(N)\psi_m$ is a function having the value zero for $n = 1$. (For $N = 1$, i.e. for $a(1) = 0$, $\psi_m(1)$ must be equal to zero by Lemma 2.)

It follows from Lemma 1, (3.2′) and the first part of the Statement that (3.11) exists in M_D iff $\sqrt[q]{N} \in Z$ and for even q $c > 0$. If this condition is satisfied, then we can write by introducing the notation μ

$$\mu = \sqrt[q]{\frac{c - \delta(N)\psi_m}{\delta(N)}} = \frac{\sqrt[q]{c - \delta(N)\psi_m}}{\delta(\sqrt[q]{N})}, \tag{3.11′}$$

$$\frac{1}{\mu} = \frac{\delta(\sqrt[q]{N})}{\sqrt[q]{c - \delta(N)\psi_m}}.$$

It is easily seen that

$$\mu \in E \quad \text{iff} \quad N = 1$$

$$\frac{1}{\mu} \in E \quad \text{for every} \quad N \in Z.$$

By taking into account (3.9), (3.10) it is obvious that the theorem holds.

III. By (3.4) we have

$$x = \frac{e^{-\int (a - a(1))}}{\sqrt[m-1]{c\delta(N) - \psi_m}}, \quad m > 2, \tag{3.12}$$

$$x = \sqrt[|m|+1]{c\delta(N) - \psi_m}\, e^{-\int (a - a(1))}, \quad m < 0. \tag{3.13}$$

By the above Corollary $k \neq N$. By our assumption (3.2‴), (3.3) is satisfied, so (3.4) holds. Since $b \in \tilde{E}_k$, so $\psi_m \in \tilde{E}_k$, and if $k \mid N$, then

$$c\delta(N) - \psi_m \in \tilde{E}_k$$

also holds. Taking into account the proof of Case I it is easily seen that the statement of Case III of Theorem 2 holds for $k \mid N$.

If $N \mid k$, then from the Corollary follows that (3.4) holds. We write for $q \in Z$

$$\sqrt[q]{c\delta(N) - \psi_m} = \sqrt[q]{\delta(N)\left[c - \delta\left(\frac{k}{N}\right)\frac{\psi_m}{\delta(k)}\right]}. \tag{3.14}$$

Obviously, $\delta\left(\frac{k}{N}\right)\frac{\psi_m}{\delta(k)} \in E$ having the value zero for $n=1$. Comparing (3.11′), (3.14) we obtain that the existence criteria of Case II hold true and

$$x \in E, \quad \text{for} \quad m < 0,$$
$$x \notin E, \quad \text{for} \quad m > 2.$$

REMARK. Our discussion can be easily extended to arbitrary positive or negative rational values of m, since for $z(k) > 0$ the formula (3.2′) can be generalized to arbitrary rational values of q in the usual way. We leave this to the reader.

REFERENCES

[1] FÉNYES, T. and KOSIK, P., The algebraic derivative and integral in the discrete operational calculus, II, *Studia Sci. Math. Hungar.* **10** (1975), 365–380. *MR* **81b:** 44017

[2] FÉNYES, T. and SZILÁRD, K., Über diskrete Mikusińskische Operatoren, die auf Grund der Dirichletschen Produktenformel erzeugt werden, *Studia Sci. Math. Hungar.* **11** (1976), 181–199. *MR* **81b:** 44014b

[3] FÉNYES, T., On a discrete nonlinear operational differential equation system based on the Dirichlet product, *Studia Sci. Math. Hungar.* **22** (1987), 471–484. *MR* **89g:** 44006

[4] GESZTELYI, E., The application of the operational calculus in the theory of numbers, *Number Theory* (Colloq., János Bolyai Math. Soc., Debrecen, 1968), North-Holland, Amsterdam, 1970, 51–104. *MR* **42** #5922

(Received June 13, 1989)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 131–134

JOIN DECOMPOSITIONS IN LOWER CONTINUOUS LATTICES

A. WALENDZIAK

1. Introduction

The results of the present paper are a continuation of those of paper [2].

Let L be a complete lattice. We say that L is lower continuous iff for every $a \in L$ and for every chain $C \subseteq L$, $a \vee \bigwedge C = \bigwedge(a \vee c : c \in C)$. A lattice is lower continuous iff its dual is upper continuous. Therefore, from Theorem 2.3 [1] it follows that the dual of an algebraic lattice is always lower continuous.

For two elements $a, b \in L$ $(a \geqq b)$ we define

$$a/b := \{x \in L : b \leqq x \leqq a\}.$$

An element $c \in a/b$ is called completely join-irreducible in a/b iff, for all $T \subseteq a/b$, $c = \bigvee T$ implies $c \in T$. We denote by $J(a/b)$ the set of all completely join-irreducible elements of a/b.

If $a \in L$, then a representation $a = \bigvee T$ with $T \subseteq J(L)$ is called a (join) decomposition of a. A decomposition $a = \bigvee T$ is irredundant if $\bigvee(T - \{t\}) \neq a$ for all $t \in T$.

In this paper we shall study infinite join decompositions of elements of lower continuous lattices.

2. The existence of decompositions

Let L be a lattice and let $\prec$ denote the covering relation in L. L is said to be weakly atomic iff for every pair of elements $a, b \in L$ with $b < a$, there exist two elements $u, v \in a/b$ such that $u \prec v$. L is called strongly dually atomic (cf. [2]) iff for every pair of elements $a, b \in L$ with $b < a$, there is an element $p \in a/b$ covered by a. Each strongly dually atomic lattice is weakly atomic.

Our first theorem is a generalization of the classical existence theorem (cf. [1], p. 43).

1980 *Mathematics Subject Classification* (1985 Revision). Primary 06B05.

Key words and phrases. Lower continuous lattices, completely join-irreducible elements, irredundant decompositions, lattices with unique irredundant decompositions.

Akadémiai Kiadó, Budapest

THEOREM 1. *If a lower continuous lattice L is weakly atomic, then every element of L has a decomposition.*

PROOF. Let a be an arbitrary element of L, and we set

$$b := \bigvee(x \in J(L) : x \leqq a).$$

Suppose now $b < a$. Since L is weakly atomic, there exist $u, v \in a/b$ such that $u \prec\!\!\!- v$. Let P be the set of all $p \in L$ with $v = u \vee p$. P is nonempty, since $v \in P$. Let C be a chain in P. By lower continuity, $u \vee \bigwedge C = \bigwedge(u \vee \vee c : c \in C) = v$. Then $\bigwedge C \in P$ and P contains a minimal element q by Zorn's Lemma. Obviously, $q \in J(L)$ and by the definition of the element b we have inequality $q \leqq b$. Hence $v = u \vee q \leqq u \vee b = u$, a contradiction. Therefore $a = \bigvee(x \in J(L) : x \leqq a)$ is a decomposition of a, and thus the proof is complete.

We say that a complete lattice L satisfies the property $(*)$ (cf. [2], p. 243) iff $a \in L$ and $b \in J(L)$ imply $a \vee b \in J(a \vee b/a)$. It is obvious that every modular lattice has this property.

Now we prove the next

THEOREM 2. *Let L be a complete lattice satisfying the condition $(*)$. If every element of L has a decomposition, then L is weakly atomic.*

PROOF. Let $a, b \in L$ with $b < a$, and let $a = \bigvee T$ be a decomposition. Since $b < a$, there is an element $t_0 \in T$ such that $t_0 \not\leqq b$. We set

$$v := t_0 \vee b \quad \text{and} \quad u := \bigvee(x \in L : b \leqq x < v)$$

(u exists, since $b < v$ and L is complete). From $(*)$ it follows that v is completely join-irreducible in v/b, and hence $u < v$. Now, by the definition of u we obtain that $u \prec\!\!\!- v$. Then L is weakly atomic.

As a consequence of Theorems 1 and 2 we get the following existence theorem:

THEOREM 3. *Let L be a lower continuous lattice satisfying $(*)$. Every element of L has a decomposition iff L is weakly atomic.*

From Theorems 8 and 10 of [2] it follows

THEOREM 4. *Let L be a lower continuous lattice satisfying $(*)$. Every element of L has an irredundant decomposition iff L is strongly dually atomic.*

3. Lattices with unique irredundant decompositions

We say that a complete lattice L has replaceable irredundant decompositions if each element of L has at least one irredundant decomposition and whenever $a = \bigvee T = \bigvee R$ are two irredundant decompositions of an element

$a \in L$, for each $t_0 \in T$ there exists $r_0 \in R$ such that $a = r_0 \vee \bigvee(T - \{t_0\})$ and this decomposition of a is irredundant. If every element of a complete lattice L has exactly one irredundant decomposition, then we say that L has unique irredundant decompositions.

G. Richter has proved the following

THEOREM 5 (cf. [2], p. 248). *A strongly dually atomic lower continuous lattice L has replaceable irredundant decompositions iff L satisfies* $(*)$.

Now we need the following

LEMMA 1. *Let L be a lower continuous strongly dually atomic lattice. Then L satisfies* $(*)$ *if L has the following property:*

$(**)$ *For every $a \in L$ and for every $x, y \in J(L)$, if $x \vee a = y \vee a$ and $x \vee y \nleqq a$, then $x = y$.*

PROOF. Suppose that L does not satisfy $(*)$. Then there is an element $a \in L$ and an element $x \in J(L)$ with $b = a \vee x \notin J(b/a)$. Hence, since L is strongly dually atomic we conclude that b/a contains two distinct dual atoms c and d. An application of lower continuity and Zorn's Lemma yields the existence of an element $y \leqq d$ which is minimal with respect to the property that $y \vee c = b$. Clearly, y is completely join-irreducible in L. Thus we have $x, y \in J(L)$, $b = x \vee c = y \vee c$ and $x \vee y \nleqq c$. Hence using $(**)$ we obtain $x = y$. Then $x \leqq d$ and consequently $b = a \vee x \leqq a \vee d = d < b$. This contradiction shows that L satisfies condition $(*)$.

Finally, we shall prove the following

THEOREM 6. *A strongly dually atomic lower continuous lattice L has unique irredundant decompositions iff L satisfies* $(**)$.

PROOF. Let us assume that L has unique irredundant decompositions but it does not satisfy $(**)$. Then there is an element $a \in L$ and there are two distinct elements $x, y \in J(L)$ such that $x \vee a = y \vee a = b$ and $x \vee y \nleqq a$. By lower continuity, there are elements $c_1, c_2 \leqq a$ which are minimal with respect to $x \vee c_1 = b$ and $y \vee c_2 = b$, respectively. Let $c_1 = \bigvee T$ and $c_2 = \bigvee R$ be irredundant decompositions of c_1 and c_2, respectively. Then

$$b = x \vee \bigvee T = y \vee \bigvee R$$

are two irredundant decompositions of b. They are also distinct, since $x \neq y$ and $x \notin R$. This contradiction proves that L has the property $(**)$.

Conversely, suppose that L satisfies $(**)$. From Lemma 1 it follows that L satisfies $(*)$, and therefore every element of L has irredundant decompositions by Theorem 4. Let $a \in L$ and $a = \vee T = \vee R$ be two irredundant decompositions of a. Pick an element $t \in T$ and we set $\bar{t} := \bigvee(T - \{t\})$. By Theorem 5 there is an element $r \in R$ such that $a = r \vee \bar{t}$. Then, $a = r \vee \bar{t} = t \vee \bar{t}$.

Moreover, $r \vee t \not\leqq \bar{t}$, because decomposition $a = \bigvee T$ is irredundant. Applying $(**)$ we conclude that $r = t$. Hence $R = T$, and thus every element of L has unique irredundant decomposition.

REFERENCES

[1] CRAWLEY, P. and DILWORTH, R. P., *Algebraic theory of lattices*, Prentice-Hall, Englewood Cliffs, N.J., 1973.

[2] RICHTER, G., The Kuroš–Ore theorem, finite and infinite decompositions, *Studia Sci. Math. Hungar.* **17** (1982), 243–250. *MR* **86c:** 06015

(Received June 18, 1989)

ZAKŁAD MATEMATYKI
WYŻSZA SZKOŁA ROLNICZO-PEDAGOGICZNA
UL. 3-GO MAJA 54
PL–08110 SIEDLCE
POLAND

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 135–144

POLYNOMIAL APPROXIMATION ON LOCALLY COMPACT ABELIAN GROUPS

R. WINKLER

Abstract

The set of all unary functions $f: G \to G$ in a locally compact abelian Hausdorff group G which can be approximated pointwise or uniformly by polynomial functions is studied.

1. Introduction

Let $\langle G, + \rangle$ be an abelian group and $\mathcal{T}^*$ a topology on the set

$$\mathcal{F}_G = G^G = \{f \mid f: G \to G\}$$

of all maps from G to G. We shall consider the set

$$(1) \qquad \mathcal{P} = \bigcup_{k \in \mathbf{Z}} \mathcal{P}_k, \qquad \mathcal{P}_k = \{f \in \mathcal{F} \mid \exists a \in G \, \forall x \in G : f(x) = a + kx\},$$

of polynomial functions on G and ask: "What does the set $\bar{\mathcal{P}}$ (topological closure of $\mathcal{P}$ with respect to $\mathcal{T}^*$) look like?" Throughout the paper by G we mean — if not specified in a different manner — a locally compact abelian Hausdorff group. There are two reasons why this class of topological groups can be investigated very effectively.

The first one: There is a very strong duality theory for locally compact abelian groups, which will be used several times.

The second one: The set $\mathcal{P}$ of polynomials is given by a handy set of normal-forms, cf. (1).

In the following we shall consider two topologies $\mathcal{T}^*$, that of uniform and that of pointwise convergence. In both cases we use the notion of a character χ on a topological group G, that is, a continuous homomorphism which maps the topological group $\langle G, + \rangle$ to the unit circle (one-dimensional torus) $\mathbf{T} := \langle \{z \in \mathbf{C} \mid |z| = 1\}, \cdot \rangle$ ($\cong \langle \mathbf{R}, + \rangle / \langle \mathbf{Z}, + \rangle$) in the complex plane with the natural topology. The set of characters on G with multiplication $\cdot$ defined pointwise and the topology of uniform convergence on compact subsets again forms a topological abelian group $\langle \Xi_G, \cdot \rangle$, the character group of G.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 22B05; Secondary 08A40.

Key words and phrases. Pointwise approximation, uniform approximation, polynomial endomorphisms, congruence compatible functions, topological algebras.

Akadémiai Kiadó, Budapest

2. Uniform approximation

Let $\mathcal{T}$ be a topology on G. It is well known that, for every neighbourhood base $\mathcal{U}$ of 0 (for the rest of this paper $\mathcal{U}$ is always used in this meaning) with respect to $\mathcal{T}$,

$$\mathcal{B}_f := \{\{g \in \mathcal{F}_G \mid g(x) - f(x) \in U \ \forall x \in G\} \mid U \in \mathcal{U}\} \tag{2}$$

is a neighbourhood base of $f \in \mathcal{F}_G$ with respect to the topology $\mathcal{T}^*$ of uniform convergence. In most cases the set $\bar{\mathcal{P}}$ turns out to be "very small". This fact is expressed by

THEOREM 1. *Let G be not totally disconnected (i.e., there are connected subsets with more than one element). Then $\mathcal{P}$ is closed in the topology of uniform convergence and is the countable union of the separated closed sets $\mathcal{P}_k = \{p = a + kx \mid a \in G\}$, each of them homeomorphic to G. Hence only (trivial) approximation of polynomials by polynomials (with the same k) is possible.*

PROOF. It suffices to prove the following four facts:

i) $\exists U \in \mathcal{U}\colon \forall k_1 \neq k_2 \in \mathbf{Z},\ a_1, a_2 \in G\colon \exists x \in G\colon$

$$(k_1 x + a_1) - (k_2 x + a_2) \notin U$$

ii) $f \in \bar{\mathcal{P}} \Rightarrow \exists k \in \mathbf{Z}\colon f \in \bar{\mathcal{P}}_k$

iii) $\forall k \in \mathbf{Z}\colon f \in \bar{\mathcal{P}}_k \Rightarrow f \in \mathcal{P}_k$

iv) $\forall k \in \mathbf{Z}\colon \phi_k\colon G \to \mathcal{P}_k,\ a \mapsto \phi_k(a) := a + kx$, is homeomorphism.

ad i): From duality theory it is known that, for G not totally disconnected, there exists a character $\chi \in \Xi_G$ which is onto, i.e., $\chi(G) = \mathbf{T}$. $U := \chi^{-1}(\{z \in \mathbf{T} \mid z \neq -1\})$ is an open neighbourhood of 0. With $k := k_1 - k_2$, $a := a_1 - a_2$ there exists a solution $x \in G$ of $\chi(x)^k = -\bar{\chi}(a)$ ($\bar{z}$ denotes the conjugate complex number of $z \in \mathbf{C}$) because χ is onto.

$\Rightarrow \chi((k_1 x + a_1) - (k_2 x + a_2)) = \chi(kx + a) = \chi(x)^k \chi(a) = -1$

$\Rightarrow (k_1 x + a_1) - (k_2 x + a_2) \notin U$.

ad ii): Let $f \in \bar{P}$, U from i). From the continuity of the group operations follows

$$\exists V \in \mathcal{U}\colon 2V \subseteq U \wedge -V = V$$

$$(2V = V + V = \{v_1 + v_2 \mid v_1, v_2 \in V\},\ -V = \{-v \mid v \in V\}).$$

$$\exists a_1 \in G,\ k_1 \in \mathbf{Z}\colon \forall x \in G\colon f(x) - k_1 x - a_1 \in V.$$

Claim: $f \in \bar{\mathcal{P}}_{k_1}$. Let $W \in \mathcal{U}$ arbitrary, w.l.o.g. $W \subseteq V$.

$$f \in \bar{\mathcal{P}} \Rightarrow$$

$$\exists a_2 \in G,\ k_2 \in \mathbf{Z}\colon \forall x \in G\colon f(x) - k_2 x - a_2 \in W \Rightarrow$$

$$(k_1 x + a_1) - (k_2 x + a_2) = (f(x) - k_2 x - a_2) - (f(x) - k_1 x - a_1) \in 2V \subseteq U.$$

Hence by i) $k_2 = k_1$, therefore indeed $f \in \bar{\mathcal{P}}_{k_1}$.

ad iii): For $f \in \bar{\mathcal{P}}_k$ we claim $\forall x \in G\colon f(x) = f(0) + kx$. Let $W \in \mathcal{U}$ arbitrary. We shall show $f(0) + kx - f(x) \in W\ \forall x \in G$: Take $V \in \mathcal{U}$ such that $2V \subseteq W$, $-V = V$.

$$f \in \bar{\mathcal{P}}_k \Rightarrow \exists a \in G\, \forall x \in G\colon a + kx - f(x) \in V \Rightarrow \forall x \in G\colon f(0) + kx - f(x) =$$
$$= (a + kx - f(x)) + (f(0) - a - k0) \in V + V \subseteq W.$$

Since $W \in \mathcal{U}$ is arbitrarily chosen and G satisfies the Hausdorff separation axiom, we have $\forall x \in G\colon f(x) = f(0) + kx$ and thus $f \in \mathcal{P}_k$.

ad iv): trivial.

If G is totally disconnected, Theorem 1 fails to be true. This can be shown by the following counterexample:

Let $G := \mathbf{Z}_2 \times \mathbf{Z}_3 \times \mathbf{Z}_5 \times \ldots$ the direct product of infinitely many different discrete cyclic groups of prime order and $\mathcal{T}$ the product topology. By Tychonoff's theorem G is a compact abelian T_2-group. Let us consider $f\colon G \to G, (a_1, a_2, \ldots, a_n, \ldots) \mapsto (1a_1, 2a_2, \ldots, na_n, \ldots)$.

Claim i): $f \in \bar{\mathcal{P}}$: Let $U \in \mathcal{U}$. According to the definition of product topology there is an $n \in \mathbf{N}$ with $\{0\} \times \ldots \times \{0\} \times \mathbf{Z}_{p_{n+1}} \times \mathbf{Z}_{p_{n+2}} \times \ldots \subseteq U$. The Chinese Remainder Theorem guarantees the existence of a solution k of the congruence system $k \equiv i \bmod p_i$, $i = 1, \ldots, n$, therefore $kx - f(x) \in U\ \forall x \in G$, hence indeed $f \in \bar{\mathcal{P}}$.

Claim ii): $f \notin \mathcal{P}$: $f \in \mathcal{P}$ and $f(0) = 0$ would imply $f = kx$ with $k \equiv i \bmod p_i$ for all primes p_i which is impossible.

3. Pointwise approximation

With the same notation as in (2) a neighbourhood base $\mathcal{B}_f$ of a function f with respect to the topology $\mathcal{T}^*$ of pointwise convergence has the form

$$\mathcal{B}_f = \{\{g \in \mathcal{F} \mid g(x) - f(x) \in U\ \forall x \in T\} \mid U \in \mathcal{U}, T \subseteq G, T \text{ finite}\}.$$

For the rest of the paper we only consider this topology. For the description of the set $\bar{\mathcal{P}}$ we also use the following notations:

$$\mathcal{F}_0 := \{f \in \mathcal{F} \mid f(0) = 0\} \ldots \text{ "normalized functions"}$$

$$\mathcal{P}_0 := \{f \in \mathcal{F} \mid \exists k \in \mathbf{Z}\, \forall x \in G\colon f(x) = kx\} \ldots \text{ "power functions"}$$

$$\mathcal{K} := \{f \in \mathcal{F} \mid \forall N\colon N \text{ closed subgroup of } G \Rightarrow f(N) \subseteq N\} \ldots$$
"maps respecting closed subgroups"

$$\underline{\mathcal{K}} := \{f \in \mathcal{F} \mid \forall n \in \mathbf{N}\, \forall N\colon N \text{ closed subgroup of } G^n$$
$$\text{and } (a_1, \ldots, a_n) \in N \Rightarrow (f(a_1), \ldots, f(a_n)) \in N\} \ldots$$
"maps respecting closed subgroups of powers"

$\mathcal{E} := \{f \in \mathcal{F} \mid \forall a, b \in G\colon f(a+b) = f(a) + f(b)\}$ ... "endomorphisms of G"

LEMMA 1. *In every topological abelian group G the following equivalence holds for every $f \in \mathcal{F}$:*

$$f \in \bar{\mathcal{P}} \Leftrightarrow g := f - f(0) \in \bar{\mathcal{P}}_0.$$

PROOF. Clear.

By Lemma 1 we may restrict our investigations to the smaller set $\bar{\mathcal{P}}_0$.

The main result, which gives the relations between the classes of functions listed above, is

THEOREM 2.

$$\mathcal{P}_0 \subseteq \bar{\mathcal{P}}_0 = \underline{\mathcal{K}} \subseteq \mathcal{E} \cap \mathcal{K} \begin{matrix} \subseteq \\ \subseteq \end{matrix} \begin{matrix} \mathcal{E} \\ \mathcal{K} \end{matrix} \begin{matrix} \subseteq \\ \subseteq \end{matrix} \mathcal{E} \cup \mathcal{K} \subseteq \mathcal{F}_0 \subseteq \mathcal{F}.$$

First we prove

LEMMA 2. *The following conditions are equivalent:*

(i) $f \in \bar{\mathcal{P}}_0$

(ii) $\forall a_1, \ldots, a_n \in G\colon (f(a_1), \ldots, f(a_n)) \in \langle (a_1, \ldots, a_n) \rangle$

(iii) $\forall a_1, \ldots, a_n \in G,\ \chi_1, \ldots, \chi_n \in \Xi_G$:

$$\chi_1(a_1) \cdot \ldots \cdot \chi_n(a_n) = 1 \Rightarrow \chi_1(f(a_1)) \cdot \ldots \cdot \chi_n(f(a_n)) = 1$$

(iv) $f \in \underline{\mathcal{K}}$.

In (ii) $\langle (a_1, \ldots, a_n) \rangle$ *denotes the topological closure of the subgroup of G^n generated by the element $(a_1, \ldots, a_n) \in G^n$.*

PROOF. (i)$\Longleftrightarrow$(ii). $f \in \bar{\mathcal{P}}_0 \Leftrightarrow \forall U \in \mathcal{U},\ a_1 \ldots, a_n \in G\ \exists k \in Z\colon f(a_i) - ka_i \in U,\ i = 1, \ldots, n \Leftrightarrow$(ii)

(ii)$\Longrightarrow$(iii). Let $\chi_i \in \Xi_G$, $i = 1, \ldots, n$ and $\chi_1(a_1) \cdot \ldots \cdot \chi_n(a_n) = 1$. Then $\chi(x_1, \ldots, x_n) := \chi_1(x_1) \cdot \ldots \cdot \chi_n(x_n)$ defines a character $\chi \in \Xi_{G^n}$. $\Rightarrow \chi^{-1}(\{1\})$ is a closed subgroup of G^n containing $(a_1, \ldots, a_n) \Rightarrow$

$$(f(a_1), \ldots, f(a_n)) \in \langle (a_1, \ldots, a_n) \rangle \subseteq \chi^{-1}(\{1\}) \Rightarrow$$
$$\Rightarrow \chi_1(f(a_1)) \cdot \ldots \cdot \chi_n(f(a_n)) = 1.$$

(iii)$\Longrightarrow$(iv). Take any $f \in \mathcal{F}$ which satisfies (iii). Let N be a closed subgroup of G^n and $(a_1, \ldots, a_n) \in N$. If $(f(a_1), \ldots, f(a_n)) \notin N$ then — as is well-known from duality theory — there exists a $\chi \in \Xi_{G^n}$ of the form $\chi((x_1, \ldots, x_n)) = \chi_1(x_1) \cdot \ldots \cdot \chi_n(x_n)$ with $\chi_i \in \Xi_G$ such that $\chi(N) = 1$ and $\chi(f(a_1), \ldots, f(a_n)) \neq 1$, which contradicts (iii).

(iv)$\Longrightarrow$(ii). $\langle (a_1, \ldots, a_n) \rangle$ is a closed subgroup of G^n, hence (iv) immediately gives $(f(a_1), \ldots, f(a_n)) \in \langle (a_1, \ldots, a_n) \rangle$.

PROOF OF THEOREM 2. $\mathcal{P}_0 \subseteq \mathcal{E} = \bar{\mathcal{E}}$ together with Lemma 2 implies $\underline{\mathcal{K}} = \bar{\mathcal{P}}_0 \subseteq \bar{\mathcal{E}} = \mathcal{E}$. The remaining relations in Theorem 2 are trivial.

COROLLARY 1 (cf. [2]). *If $|G| > 2$ then there is a function $f: G \to G$ which cannot be approximated pointwise by polynomial functions.*

PROOF. By Lemma 1 and Theorem 2 it suffices to find an $f \notin \mathcal{E}$ with $f(0) = 0$. But every map with $f(0) = 0$, $f(x) = a \neq 0$ for all $x \neq 0$ does this job.

COROLLARY 2. *$f \in \bar{\mathcal{P}}$ (i.e. "f can be approximated by polynomials pointwise") if and only if f "respects closed congruences in the powers G^n", i.e., for every $n \in \mathbf{N}$, every closed subgroup $N \subseteqq G^n$ and all $a_i, b_i \in G$, $i = 1, \ldots, n$ f satisfies the implication*

$$(a_1, \ldots, a_n) - (b_1, \ldots, b_n) \in N \Rightarrow$$
$$(f(a_1), \ldots, f(a_n)) - (f(b_1), \ldots, f(b_n)) \in N.$$

PROOF. Clear.

Now we are going to look at some special groups.

THEOREM 3. *In the case of a finite abelian group with discrete topology the identity $\mathcal{P}_0 = \bar{\mathcal{P}}_0 = \underline{\mathcal{K}} = \mathcal{E} \cap \mathcal{K}$ holds.*

PROOF. It suffices to show $\mathcal{E} \cap \mathcal{K} \subseteqq \mathcal{P}_0$. G has a representation $G = \mathbf{Z}_{m_1} \times \times \ldots \times \mathbf{Z}_{m_n}$ as a direct product of cyclic groups of orders $m_n | m_{n-1} | \ldots | m_1$, forming a chain of divisors. Let $f \in \mathcal{E} \cap \mathcal{K}$. We have to find a $k \in \mathbf{Z}$ with $f(x) = kx \ \forall x \in G$.

$$N_i := \{(0, \ldots, 0, l, 0, \ldots, 0) \mid l \in \mathbf{Z}_{m_i}\}$$

is a (closed) subgroup, hence $f \in \mathcal{K}$ implies

$$f((0, \ldots, 0, 1, 0, \ldots, 0)) = (0, \ldots, 0, k_i, 0, \ldots, 0)$$

with $0 \leqq k_i < m_i$. Furthermore

$$M_i := \{(0, \ldots, 0, l_i, l_{i+1}, 0, \ldots, 0) \mid l_i \equiv l_{i+1} \bmod m_{i+1}\}$$

is a closed subgroup, too, hence (using $f \in \mathcal{E} \cap \mathcal{K}$) one gets immediately

$$(0, \ldots, 0, k_i, k_{i+1}, 0, \ldots, 0) = f((0, \ldots, 0, 1, 1, 0, \ldots, 0)) \in M_i,$$

therefore $k_i \equiv k_{i+1} \bmod m_{i+1}$ for $i = 1, \ldots, n-1$. Now obviously we have $k_i \equiv k := k_1 \bmod m_i$ for all $i = 1, \ldots, n$, thus

$$f(a_1, \ldots, a_n) = \sum_{i=1}^{n} a_i f(0, \ldots, 0, 1, 0, \ldots, 0) =$$

$$= \sum_{i=1}^{n} a_i (0, \ldots, 0, k, 0, \ldots, 0) = k(a_1, \ldots, a_n)$$

for all $(a_1, \ldots, a_n) \in G$, the desired result.

LEMMA 3. *Let $(G_i)_{i\in I}$ be a family of abelian Hausdorff groups (locally compact or not). Consider the direct product $G = \prod_{i\in I} G_i$. Then every $f \in \in \mathcal{K}_G \cap \mathcal{E}_G$ has the form*

$$f((x_i)_{i\in I}) = (f_i(x_i))_{i\in I}$$

with $f_i \in \mathcal{K}_{G_i} \cap \mathcal{E}_{G_i}$. For the case that there exists a topological isomorphism $\phi\colon G_{i_1} \to G_{i_2}$ between two factors $G_{i_1} \cong G_{i_2}$ of the product we even have $\phi \circ f_{i_1} = f_{i_2} \circ \phi$.

PROOF. For every $i_0 \in I$, $x \in G_{i_0}$ let $f_{i_0} : G_{i_0} \to G_{i_0}$ be defined by $f_{i_0}(x) := \pi_{i_0}(f((x_i)_{i\in I}))$, with $x_{i_0} = x$ and $x_i = 0$ for $i \neq i_0$, where $\pi_{i_0}\colon G \to G_{i_0}$, $(x_i)_{i\in I} \mapsto x_{i_0}$, is the projection to the i_0-th coordinate. We show

$$\pi_{i_0}(f((x_i)_{i\in I})) = f_{i_0}(\pi_{i_0}((x_i)_{i\in I})) \quad \forall (x_i)_{i\in I} \in G.$$

For an arbitrary $(x_i)_{i\in I} \in G$ let us consider the element $(y_i)_{i\in I}$ defined by $y_{i_0} := x_{i_0}$ and $y_i := 0$ for $i \neq i_0$ and define the closed subgroup $N_{i_0} := \{(x_i)_{i\in I} \mid x_{i_0} = 0\}$. $(x_i)_{i\in I} - (y_i)_{i\in I} \in N_{i_0}$ and $f \in \mathcal{K}_G \cap \mathcal{E}_G$ imply $f((x_i)_{i\in I}) - f((y_i)_{i\in I}) \in N_{i_0}$, hence $\pi_{i_0}(f((x_i)_{i\in I})) = \pi_{i_0}(f((y_i)_{i\in I})) = f_{i_0}(x_{i_0})$. Of course indeed $f_{i_0} \in \mathcal{K}_{G_{i_0}} \cap \mathcal{E}_{G_{i_0}}$, which proves the first assertion. The second one follows by considering the closed subgroup

$$N_{i_1,i_2} := \{(x_i)_{i\in I} \mid \phi(x_{i_1}) = x_{i_2} \wedge i \notin \{i_1, i_2\} \Rightarrow x_i = 0\}$$

and using $f \in \mathcal{K}_G$.

In the following $\mathbf{R}$ and $\mathbf{Z}$ denote the additive groups of the real resp. integral numbers, $\{z \in \mathbf{C} \mid |z| = 1\}$ the multiplicative group of complex numbers on the unit circle. Furthermore we define

$$\mathbf{T} := \mathbf{R}/\mathbf{Z} \cong \{z \in \mathbf{C} \mid |z| = 1\} \ldots \text{the one-dimensional torus}$$

and

$$\mathbf{T}^n := \mathbf{R}^n/\mathbf{Z}^n \cong \mathbf{T} \times \ldots \times \mathbf{T} \ (n \text{ times}) \ldots \text{ the } n\text{-dimensional torus.}$$

THEOREM 4. *For $G = \mathbf{T}^n$, $n \in \mathbf{N}$ we have*

$$\bar{\mathcal{P}}_0 = \underline{\mathcal{K}} = \mathcal{E} \cap \mathcal{K}.$$

PROOF. By Theorem 2 we have to show $\mathcal{E} \cap \mathcal{K} \subseteq \bar{\mathcal{P}}_0$. We use the characterization (iii) of Lemma 2. Let

$$\chi_1(X_1) \cdot \ldots \cdot \chi_k(X_k) = 1,$$

$$X_i = (x_1^{(i)}, \ldots, x_n^{(i)}) \in \mathbf{T}^n, \ x_j^{(i)} \in \mathbf{T},$$

χ_i characters of $\mathbf{T}^n$ and $f \in \mathcal{E} \cap \mathcal{K}$. Lemma 3 gives $f(x_1, \ldots, x_n) = (\bar{f}(x_1), \ldots, \bar{f}(x_n))$ with $\bar{f} \in \mathcal{E}_{\mathbf{T}} \cap \mathcal{K}_{\mathbf{T}}$. We know (see [1]) that every character χ_i of $\mathbf{T}^n$ has the form

$$\chi_i(x_1, \ldots, x_n) = e(k_1^{(i)} x_1 + \ldots + k_n^{(i)} x_n)$$

with $e(x) = e^{2\pi i x}$. Hence we get

$$1 = e\Big(\sum_{i=1}^{k} \sum_{j=1}^{n} k_j^{(i)} x_j^{(i)}\Big),$$

which implies

$$\sum_{i=1}^{k} \sum_{j=1}^{n} k_j^{(i)} x_j^{(i)} \in \mathbf{Z},$$

thus for $\bar{f} \in \mathcal{E}_{\mathbf{T}}$

$$\bar{f}\Big(\sum_{i=1}^{k} \sum_{j=1}^{n} k_j^{(i)} x_j^{(i)}\Big) \in \mathbf{Z}$$

and

$$1 = e\left(\bar{f}\Big(\sum_{i=1}^{k} \sum_{j=1}^{n} k_j^{(i)} x_j^{(i)}\Big)\right) = e\Big(\sum_{i=1}^{k} \sum_{j=1}^{n} k_j^{(i)} \bar{f}(x_j^{(i)})\Big) =$$
$$= \chi_1(f(X_1)) \cdot \ldots \cdot \chi_k(f(X_k)).$$

By Lemma 2 (iii) the proof of Theorem 4 is now complete.

THEOREM 5. *In* $\mathbf{T}$ *furthermore the inclusion* $\mathcal{E} \subseteq \underline{\mathcal{K}}$ *holds and therefore we have* $\bar{\mathcal{P}}_0 = \underline{\mathcal{K}} = \mathcal{E} \cap \mathcal{K} = \mathcal{E}$, *the inclusion* $\mathcal{P}_0 \subset \bar{\mathcal{P}}_0$ *is strict.*

PROOF. With

$$\chi_j(x) = e(k_j x) = e^{2\pi i k_j x}, \quad f \in \mathcal{E} \text{ and } \chi_1(x_1) \cdot \ldots \cdot \chi_n(x_n) = 1$$

we get

$$1 = e\Big(\sum_{j=1}^{n} k_j x_j\Big), \quad \sum_{j=1}^{n} k_j x_j \in \mathbf{Z}$$

and

$$\sum_{j=1}^{n} k_j f(x_j) = f\Big(\sum_{j=1}^{n} k_j x_j\Big) \in \mathbf{Z},$$

thus $\chi_1(f(x_1)) \cdot \ldots \cdot \chi_n(f(x_n)) = 1$. Lemma 2 gives $\mathcal{E} \subseteq \underline{\mathcal{K}}$ and hence by Theorem 2 $\bar{\mathcal{P}}_0 = \underline{\mathcal{K}} = \mathcal{E} \cap \mathcal{K} = \mathcal{E}$.

In order to prove the second statement we mention that by $\bar{\mathcal{P}}_0 = \mathcal{E}$ our set $\bar{\mathcal{P}}_0$ is just the dual group of the discrete torus $\mathbf{T}_d$, which is different from $\mathcal{P}_0$, cf. [1], pages 405–406.

THEOREM 6. *For $G = \mathbf{R}^n$ (n may also denote an infinite cardinality) the equalities $\mathcal{P}_0 = \bar{\mathcal{P}}_0 = \underline{\mathcal{K}} = \mathcal{E} \cap \mathcal{K}$ are valid.*

PROOF. First we prove the assertion for $n = 1$. Let $f \in \mathcal{E} \cap \mathcal{K}$. For every $a \in \mathbf{R}$ $N_a := \{ka \mid k \in \mathbf{Z}\}$ is a closed subgroup. $f \in \mathcal{K}$, hence $f(a) = k_a a$, $k_a \in \mathbf{Z}$. We have to show $k_a = k_b$ for $a \neq b$.

First case: a, b independent over $\mathbf{Q}$ (or equivalently over $\mathbf{Z}$):

$$k_{a+b}a + k_{a+b}b = k_{a+b}(a+b) = f(a+b) = f(a) + f(b) = k_a a + k_b b,$$

hence $(k_{a+b} - k_a)a + (k_{a+b} - k_b)b = 0$, which implies $k_{a+b} - k_a = k_{a+b} - k_b = 0$ and $k_a = k_{a+b} = k_b$.

Second case: a, b dependent. There exists a real number c such that c, a and c, b are independent. Then the first case gives $k_a = k_c = k_b$. Now we have found an integer $k \in \mathbf{Z}$ such that $f(x) = kx$ $\forall x \in \mathbf{R}$. The generalization from $\mathbf{R}$ to $\mathbf{R}^n$ now is an immediate consequence of Lemma 3.

The interesting yet unsolved question remains if $\underline{\mathcal{K}} = \mathcal{E} \cap \mathcal{K}$ holds in the general case. We are only able to prove

THEOREM 7. *The inclusions $\mathcal{P}_0 \subsetneqq \bar{\mathcal{P}}_0$, $\mathcal{E} \cap \mathcal{K} \subsetneqq \mathcal{E}$, $\mathcal{E} \cap \mathcal{K} \subsetneqq \mathcal{K}$, $\mathcal{E} \subsetneqq \mathcal{E} \cup \mathcal{K}$, $\mathcal{K} \subsetneqq \mathcal{E} \cup \mathcal{K}$, $\mathcal{E} \cup \mathcal{K} \subsetneqq \mathcal{F}_0$, and $\mathcal{F}_0 \subsetneqq \mathcal{F}$ of Theorem 2 (in general) cannot be replaced by equalities.*

PROOF. $\mathcal{P}_0 \neq \bar{\mathcal{P}}_0$: cf. Theorem 5.

$\mathcal{E} \cap \mathcal{K} \neq \mathcal{E}$ and hence $\mathcal{K} \neq \mathcal{E} \cup \mathcal{K}$: Take $G = G_1 \times G_1$, $f(a, b) = (b, a)$, $|G_1| > 1$.

$\mathcal{E} \cap \mathcal{K} \neq \mathcal{K}$ and hence $\mathcal{E} \neq \mathcal{K} \cup \mathcal{E}$: $G = \mathbf{Z}_p$ (p prime) is simple, hence $\mathcal{K} = \mathcal{F}_0$ but certainly $\mathcal{E} \neq \mathcal{F}_0$ for $p > 2$.

$\mathcal{E} \cup \mathcal{K} \neq \mathcal{F}_0$; Take $G = \mathbf{Z}_4$, $f\colon 0 \mapsto 0, 1, 2, 3 \mapsto 1$.

$\mathcal{F}_0 \neq \mathcal{F}$: Take $|G| > 1$, $f(0) = a \neq 0$.

REMARK. Although polynomials are continuous, the closure $\bar{\mathcal{P}}_0$ may contain functions that are not. Examples are the functions $f \in \bar{\mathcal{P}}_0 \setminus \mathcal{P}_0$ given by the second statement of Theorem 5.

REMARK. Although $\mathcal{P}_0$ is countable, for $f \in \bar{\mathcal{P}}_0$ it is not necessary that there exists a sequence $(p_n)_{n \in \mathbf{N}}$ of power functions such that $p_n(x) = k_n x \to$ $\to f(x)$ pointwise. It is even possible to prove

THEOREM 8. *Let G be compact, not totally disconnected and $(k_n x)_{n \in \mathbf{N}}$ an arbitrary sequence of power functions, k_n pairwise different, then the set M of all points x where $(k_n x)_{n \in \mathbf{N}}$ converges has Haar measure $\mu(M) = 0$.*

PROOF. First we prove the assertion for the case $G = \mathbf{T}$. We use the following "norm" $\|\cdot\|$ on $\mathbf{R}$ describing the topology on $\mathbf{T}$: $\|x\| := \min_{k \in \mathbf{Z}} |x - k|$.

With

$$M_{\varepsilon,n_1,n_2} := \{x \in \mathbf{T} \mid \|(k_{n_1} - k_{n_2})x\| < \varepsilon\}$$

and

$$M_{\varepsilon,N} := \bigcap_{n_1=N}^{\infty} \bigcap_{n_2=N}^{\infty} M_{\varepsilon,n_1,n_2}$$

we have

$$M = \{x \in \mathbf{T} \mid (k_n x)_{n\in\mathbf{N}} \text{ converges in } \mathbf{T}\} =$$
$$= \{x \in \mathbf{T} \mid \forall \varepsilon > 0\ \exists N\ \forall n_1, n_2 \geqq N\ \|(k_{n_1} - k_{n_2})x\| < \varepsilon\} =$$
$$= \bigcap_{\varepsilon>0} \bigcup_{N=1}^{\infty} M_{\varepsilon,N}$$

and $\mu(M_{\varepsilon,n_1,n_2}) \leqq 2\varepsilon$. Hence $\mu(M_{\varepsilon,N}) \leqq 2\varepsilon\ \forall N, \varepsilon > 0$ and therefore $((M_{\varepsilon,N})_{N\in\mathbf{N}}$ is a monotonous sequence of sets)

$$\mu\Big(\bigcup_{N=1}^{\infty} M_{\varepsilon,N}\Big) = \lim_{N\to\infty} \mu(M_{\varepsilon,N}) \leqq 2\varepsilon \quad \forall \varepsilon > 0$$

and

$$\mu(M) = \mu\Big(\bigcap_{\varepsilon>0} \bigcup_{N=1}^{\infty} M_{\varepsilon,N}\Big) = 0.$$

The general case can be treated in the following manner: Let us consider the sets

$$M = \{x \in G \mid (k_n x)_{n\in\mathbf{N}} \text{ converges in } G\}$$

and

$$M' = \{y \in \mathbf{T} \mid (k_n y)_{n\in\mathbf{N}} \text{ converges in } \mathbf{T}\}.$$

It is easily checked that they form subgroups of G resp. $\mathbf{T}$. Because G is not totally disconnected there is a character χ which is onto (duality theory), hence $|G/\chi^{-1}(M')| = |\mathbf{T}/M'|$ and by continuity $M \subseteqq \chi^{-1}(M')$. By the first part $\mu_{\mathbf{T}}(M') = 0$, hence by translation invariance and σ-additivity of $\mu_{\mathbf{T}}$ $\mathbf{T}/M'$ is infinite (even uncountable), therefore also $G/\chi^{-1}(M)$ is infinite and by the same argument $\mu(M) \leqq \mu(\chi^{-1}(M')) = 0$.

Most results in this paper said that, in a certain sense, there are only few functions which can be approximated by polynomials. A converse statement is

THEOREM 9. *Let G be connected and compact with a countable topological base. Then for every $n \in \mathbf{N}$ there is a set $T \subseteqq G^n$ which has Haar measure*

$\mu(T) = 0$ and is meager (i.e., countable union of sets whose closure has empty interior) such that for every $(a_1, \dots, a_n) \in G^n - T$ every $f: G \to G$ can be approximated at the points $a_1, \dots, a_n$ by polynomial functions.

PROOF. As in Lemma 2 one sees that approximation of f by power functions only is impossible if the implication

$$\chi(a_1, \dots, a_n) = 1 \Rightarrow \chi(f(a_1), \dots, f(a_n)) = 1$$

fails for a character χ of G^n. This can occur only if the assumption is satisfied. Therefore the set T of all $(a_1, \dots, a_n) \in G^n$ such that not every function can be approximated at the points $a_1, \dots, a_n$ satisfies

$$T \subseteq \bigcup \{\chi^{-1}(\{1\}) \mid \chi \in \Xi_{G^n},\ \chi \not\equiv 1\}. \tag{3}$$

First we investigate $\mu(T)$: Every nontrivial character of the connected group G^n is onto (duality theory), the Haar measure μ on the compact group G^n satisfies $\mu(G^n) = 1$ and is translation invariant, which implies ($\mathbf{T}$ is infinite) for every $z \in \mathbf{T}$

$$\mu(\chi^{-1}(\{z\})) = \mu(\chi^{-1}(\{1\})) = 0 \tag{4}$$

for every $\chi \in \Xi_{G^n}$. With G, G^n has countable base too, therefore only countably many different characters (duality theory), thus (3), (4) and σ-additivity of μ give $\mu(T) = 0$.

For showing that T is meager it suffices to show that for every nontrivial character χ the closed subgroup $N = \chi^{-1}(\{1\})$ has empty interior N^o. Suppose $N^o \neq \phi$. First we claim that N^o is a subgroup of G: $x, y \in N^o$ implies the existence of an open neighbourhood U of y such that $U \subseteq N$. $\{x - z \mid z \in U\} \subseteq N$ is an open neighbourhood of $x - y$ contained in N, hence $x - y \in N^o$. Thus indeed N^o is a subgroup. The factor group G/N^o consists of an infinite number of disjoint open classes, which contradicts compactness. Thus the proof of Theorem 9 is complete.

REFERENCES

[1] HEWITT, E. and ROSS, K. A., *Abstract harmonic analysis*, Vol. I: Structure of topological groups. Integration theory, group representations, Die Grundlehren der mathematischen Wissenschaften, Bd. 115, Academic Press, New York; Springer-Verlag, Berlin–Göttingen–Heidelberg, 1963. *MR* **28** #158

[2] KOWOL, G., Approximation durch Polynomfunktionen auf universellen Algebren, *Monatsh. Math.* **93** (1982), 15–32. *MR* **83e**: 08007

(Received June 26, 1989)

INSTITUT FÜR ALGEBRA UND DISKRETE MATHEMATIK
TECHNISCHE UNIVERSITÄT WIEN
WIEDNER HAUPTSTRASSE 8–10
A–1040 WIEN
AUSTRIA

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 145–158

NONUNIFORM CONVERGENCE RATES IN THE CENTRAL LIMIT THEOREM FOR MARTINGALES

K. JOOS

Abstract

Nonuniform convergence rates in the central limit theorem for martingale difference arrays are derived. The main result is for variables with finite moments of order $2+2\delta$, $\delta > 0$. Consequences are two results where only second moments are assumed finite. In addition examples are constructed which demonstrate the optimality of the results.

1. Introduction and results

Let $(X_{n,i}, \mathcal{F}_{n,i}, 1 \leqq i \leqq k(n), n \in \mathbb{N})$ be a martingale difference array (mda). According to the well-known central limit theorem of Brown [1] the conditions

$$\sum_{i=1}^{k(n)} \mathsf{E}(X_{n,i}^2/\mathcal{F}_{n,i-1}) \longrightarrow 1 \quad \text{in prob.} \tag{1.1}$$

and

$$\sum_{i=1}^{k(n)} \mathsf{E}(X_{n,i}^2 I(|X_{n,i}| > \varepsilon)/\mathcal{F}_{n,i-1}) \longrightarrow 0 \quad \text{in prob. } \forall \varepsilon > 0 \tag{1.2}$$

are sufficient for the validity of a CLT. With the aim to derive a convergence rate under weak conditions some authors expressed estimates in the moment terms

$$L_{n,2\delta} := \sum_{i=1}^{k(n)} \mathsf{E}(|X_{n,i}|^{2+2\delta}) \text{ and } N_{n,2\delta} := \mathsf{E}\Big(\Big|\sum_{i=1}^{k(n)} \mathsf{E}(X_{n,i}^2/\mathcal{F}_{n,i-1}) - 1\Big|^{1+\delta}\Big).$$

If $L_{n,2\delta} + N_{n,2\delta} \leqq 1$ for some $\delta > 0$, then there exists a constant $0 < C_\delta < \infty$, which depends only on δ, such that for all $x \in \mathbf{R}$

$$\Big|\mathsf{P}\Big(\sum_{i=1}^{k(n)} nX_{n,i} \leqq x\Big) - \Phi(x)\Big| \leqq C_\delta (L_{n,2\delta} + N_{n,2\delta})^{1/(3+2\delta)} (1 + |x|^{2+2\delta})^{-1}. \tag{1.3}$$

1980 *Mathematics Subject Classification.* Primary 60F05; Secondary 60G42.

Key words and phrases. Martingale central limit theorem, rate of convergence, nonuniform bounds.

Akadémiai Kiadó, Budapest

(See [4], Theorem 1.) The exponent $1/(3+2\delta)$ in the uniform part $L_{n,2\delta} + N_{n,2\delta}$ is exact: there is in Häusler [3] a martingale difference array $(X_{n,i}, \mathcal{F}_{n,i}, 1 \leqq i \leqq n, n \in \mathbb{N})$ with

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}} \left| \mathsf{P}\Big(\sum_{i=1}^{n} X_{n,i} \leqq x\Big) - \Phi(x) \right| (L_{n,2\delta} + N_{n,2\delta})^{-1/(3+2\delta)} > 0.$$

Furthermore (1.3) is exact w.r.t. the nonuniform part: In the case of independent $X_{n1}, \ldots, X_{nn}$ it is well-known that the term $(1+|x|^{2+2\delta})^{-1}$ is sharp for large $|x|$ if the variables have moments of order $2+2\delta$, see i.e. Remark 1 in [5]. Nonuniform bounds like (1.3) provide for example rates of convergence of moments and L_p-norms in the CLT, whereas these do not follow from uniform bounds. From now on let $(X_i, \mathcal{F}_i, 1 \leqq i \leqq n)$ be a martingale difference sequence (mds for short) and

$$L_{n,2\delta} := \sum_{i=1}^{n} \mathsf{E}(|X_i|^{2+2\delta}), \quad \delta > 0,$$

$$N_{n,2\delta} := \mathsf{E}\Big(\Big| \sum_{i=1}^{n} \mathsf{E}(X_i^2/\mathcal{F}_{i-1}) - 1 \Big|^{1+\delta} \Big), \quad \delta \geqq 0,$$

$$N_{n,*} := \Big\| \sum_{i=1}^{n} \mathsf{E}(X_i^2/\mathcal{F}_{i-1}) - 1 \Big\|_{\infty}.$$

Our main result is the following Theorem, which generalizes (1.3). There we drop the restriction that the Ljapunov term $L_{n,2\delta}$ and the norming term $N_{n,2\delta}$ have the same parameter δ.

THEOREM. *Let $\delta > 0$ and $L_{n,2\delta} \leqq 1/2$.*

(i) *Let $0 \leqq r < \infty$ and $N_{n,2r} \leqq 1/2$. There exists a constant $0 < C_{\delta,r} < \infty$, which depends only on δ and r, such that for all $x \in \mathbb{R}$*

$$\Big| \mathsf{P}\Big(\sum_{i=1}^{n} X_i \leqq x\Big) - \Phi(x) \Big| \leqq C_{\delta,r} \Big(L_{n,2\delta}^{1/(3+2\delta)} + N_{n,2r}^{1/(3+2r)} \Big) (1+|x|^{2+2s})^{-1},$$

where $s := \min\{\delta, r\}$.

(ii) *Let $N_{n,*} \leqq 1/2$. There exists a constant $0 < C_\delta < \infty$, which depends only on δ, such that for all $x \in \mathbb{R}$*

$$\Big| \mathsf{P}\Big(\sum_{i=1}^{n} X_i \leqq x\Big) - \Phi(x) \Big| \leqq C_\delta \Big(L_{n,2\delta}^{1/(3+2\delta)} + N_{n,*}^{1/2} \Big) (1+|x|^{2+2\delta})^{-1}.$$

Since $N_{n,2r}^{1/(3+2r)} \nearrow N_{n,*}^{1/2}$ for $r \to \infty$, it may be expected that (ii) follows directly from (i). But we need some modification for the proof of (ii) because

in (i) we use the Burkholder inequality, and therefore the constant $C_{\delta,r}$ goes to ∞.

The following example is a concrete situation, where the Theorem gives a better estimate than (1.3), that is, we have a mds $(X_{n,i}, \mathcal{F}_{n,i}, 1 \leqq i \leqq n, n \in \mathbb{N})$ with

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}} \left| \mathsf{P}\Big(\sum_{i=1}^{n} X_{n,i} \leqq x\Big) - \Phi(x) \right| (L_{n,2\delta} + N_{n,2\delta})^{-1/(3+2\delta)} = 0$$

and

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}} \left| \mathsf{P}\Big(\sum_{i=1}^{n} X_{n,i} \leqq x\Big) - \Phi(x) \right| \Big(L_{n,2\delta}^{1/(3+2\delta)} + N_{n,*}^{1/2} \Big)^{-1} > 0.$$

EXAMPLE 1. Let $n > 2$ and $\alpha_n = 1/\ln(n)$. Let $X_{n,1}, \ldots, X_{n,n-1}, Y_n$ be independent r.v. with

$$\mathcal{L}\{X_{n,i}\} = N(0, (1-\alpha_n^2)/(n-1)), \quad 1 \leqq i \leqq n-1, \quad \mathsf{P}(Y_n = \pm\alpha_n) = 1/2.$$

Let $N_{n-1} = \sum_{i=1}^{n-1} X_{n,i}$ and $X_{n,n} = Y_n I(N_{n-1} \in [0, \alpha_n])$. Then we have for all $\delta > 0$

$$L_{n,2\delta}^{1/(3+2\delta)} \sim C_\delta \alpha_n,$$

$$N_{n,2\delta}^{1/(3+2\delta)} \sim C_\delta \alpha_n^{(2+2\delta)/(3+2\delta)},$$

$$N_{n,*}^{1/2} = \alpha_n,$$

and because

$$\mathsf{P}\Big(\sum_{i=1}^{n} X_{n,i} \leqq 0\Big) = \mathsf{P}(N_{n-1} \leqq 0) + \frac{1}{2}\mathsf{P}(0 \leqq N_{n-1} \leqq \alpha_n),$$

we have

$$\left| \mathsf{P}\Big(\sum_{i=1}^{n} X_{n,i} \leqq 0\Big) - \Phi(0) \right| \geqq C\alpha_n. \qquad \square$$

As consequences of the Theorem we obtain the following results, where we only assume that the second moments are finite. Motivated by the Lindeberg condition (1.2) we define for $\beta > 0$

$$L(n,\beta) := \sum_{i=1}^{n} \mathsf{E}(X_i^2 I(|X_i| > \beta)).$$

COROLLARY 1. *Let $\beta > 0$, $\delta > 0$ and*

$$N_{n,0} < 1/8, \quad L(n,\beta) < 1/8, \quad \sum_{i=1}^{n} \mathsf{E}(|X_i|^{2+2\delta} I(|X_i| \leqq \beta)) < 2^{-3-2\delta}.$$

Then there exists a constant $0 < C_\delta < \infty$, *which depends only on* δ, *such that for all* $x \in \mathbb{R}$

$$\left|\mathsf{P}\Big(\sum_{i=1}^{n} X_i \leqq x\Big) - \Phi(x)\right| \leqq$$

$$\leqq C\left\{\Big[\sum_{i=1}^{n} \mathsf{E}(|X_i|^{2+2\delta} I(|X_i| \leqq \beta))\Big]^{1/(3+2\delta)} + L(n,\beta)^{1/3} + N_{n,0}^{1/3}\right\}(1+x^2)^{-1}.$$

In our next result we use the same Lindeberg term as Móri [6], who showed the uniform version of Corollary 2. Let

$$W_n := \int_0^1 \sum_{i=1}^{n} \mathsf{E}(X_i^2 I(|X_i| > \varepsilon))\,d\varepsilon = \int_0^1 L(n,\varepsilon)\,d\varepsilon.$$

Then we obtain as a consequence of Corollary 1

COROLLARY 2. *Let* $W_n < 1/16$ *and* $N_{n,0} < 1/8$. *There exists a constant* $0 < C < \infty$, *such that for all* $x \in \mathbb{R}$

$$\left|\mathsf{P}\Big(\sum_{i=1}^{n} X_i \leqq x\Big) - \Phi(x)\right| \leqq C\left[W_n^{1/4} + N_{n,0}^{1/3}\right]\frac{1}{1+x^2}.$$

The bound in Corollary 2 is sharp in all terms. Obviously it is exact w.r.t. the nonuniform part. That it is also exact in the uniform part we see in the next two examples. Example 2 demonstrates that the exponent $1/4$ in the Lindeberg term W_n (and moreover $1/(3+2\delta)$ in the Ljapunov term $L_{n,2\delta}$ in the Theorem) cannot be improved, even in the special case that the sum of the conditional variances $\sum_{i=1}^{n} \mathsf{E}(X_i^2/\mathcal{F}_{i-1})$ is equal to 1. Finally Example 3 shows that the exponent $1/3$ in the norming term $N_{n,0}$ is sharp.

EXAMPLE 2. Let $n > 2$ and $\alpha_n := 1/\ln(n)$. $X_{n,1}, \ldots, X_{n,n-1}, Y_{n,n}, \ldots, Y_{n,2n}$ are independent r.v. with $\mathcal{L}\{X_{n,i}\} = N(0,(1-\alpha_n^2)/(n-1))$, $1 \leqq i \leqq n-1$, $\mathsf{P}(Y_{n,n} = \pm\alpha_n) = 1/2$ and $\mathcal{L}\{Y_{n,i}\} = N(0,\alpha_n^2/n)$, $n+1 \leqq i \leqq 2n$. Let $N_{n-1} := \sum_{i=1}^{n-1} X_{n,i}$, $X_{n,n} := Y_{n,n} I(N_{n-1} \in [0,\alpha_n])$ and $X_{n,i} := Y_{n,i} I(N_{n-1} \notin [0,\alpha_n])$ for $n+1 \leqq i \leqq 2n$. Then $\sum_{i=1}^{2n} \mathsf{E}(X_{n,i}^2/\mathcal{F}_{n,i-1}) = 1$. With

$$\mathsf{E}(|X_{n,i}|^{2+2\delta})\begin{cases} \leqq C_\delta n^{-1-\delta} & \text{for } i \neq n \\ = \alpha_n^{2+2\delta}\mathsf{P}(N_{n-1} \in [0,\alpha_n]) \sim C_\delta \alpha_n^{3+2\delta} & \text{for } i = n \end{cases}$$

we get

$$L_{n,2\delta}^{1/(3+2\delta)} \sim \alpha_n$$

and

$$W_n \leqq Cn^{-1/4} + \int_0^1 \mathsf{E}(X_{n,n}^2 I(|X_{n,n}| > y))dy =$$

$$= Cn^{-1/4} + \alpha_n^2 \int_0^{\alpha_n} \mathsf{E}(I(N_{n-1} \in [0, \alpha_n]))dy \leqq C\alpha_n^4$$

implies

$$W_n^{1/4} \sim \alpha_n.$$

$$\mathsf{P}\Big(\sum_{i=1}^{n} X_{n,i} \leqq 0\Big) - \Phi(0) =$$

$$= \int_0^{\alpha_n} \Big[\frac{1}{2} - \mathsf{P}\Big(\sum_{i=n+1}^{2n} Y_{n,i} \leqq -x\Big)\Big] \mathsf{P}(N_{n-1} \in dx) =$$

$$= \int_0^{\alpha_n} \Big[\frac{1}{2} - \Phi(-x\alpha_n^{-1})\Big] \mathsf{P}(N_{n-1} \in dx)$$

and so we have

$$\Big|\mathsf{P}\Big(\sum_{i=1}^{2n} X_{n,i} \leqq 0\Big) - \Phi(0)\Big| \geqq$$

$$\geqq \Big[\Phi(0) - \Phi\Big(-\frac{1}{2}\Big)\Big] \mathsf{P}\Big(N_{n-1} \in \Big[\frac{1}{2}\alpha_n, \alpha_n\Big]\Big) \geqq C_{\alpha_n}.$$

So we have shown that the convergence rate for $\sum_{i=1}^{2n} X_{n,i}$ is α_n. □

EXAMPLE 3. Let $n > 2$ and $\alpha_n := 1/\ln(n)$. $X_{n,1}, \ldots, X_{n,n}, Y_{n,n+1}, \ldots, Y_{n,2n}$ are independent r.v. with $\mathcal{L}\{X_{n,i}\} = N(0, n^{-1})$, $1 \leqq i \leqq n$, $\mathcal{L}\{Y_{n,i}\} = N(0, \alpha_n^2 n^{-1})$, $n+1 \leqq i \leqq 2n$. Let $N_n := \sum_{i=1}^{n} X_{n,i}$ and $X_{n,i} := Y_{n,i}(N_n \in [0, \alpha_n])$ for $n+1 \leqq i \leqq 2n$. Then we have

$$\sum_{i=1}^{2n} \mathsf{E}(X_{n,i}^2 / \mathcal{F}_{n,i-1}) = (1 + \alpha_n^2) I(N_n \in [0, \alpha_n]) + I(N_n \notin [0, \alpha_n])$$

and therefore $N_{n,0}^{1/3} \sim C\alpha_n$. Furthermore $W_n^{1/4} \sim Cn^{-1/8}$. Now consider the convergence rate: Taking into account that $\mathcal{L}\{N_n\} = N(0,1)$ and $\mathcal{L}\Big\{\sum_{i=n+1}^{2n} Y_{n,i}\Big\} = N(0,\alpha_n^2)$, we obtain

$$\mathsf{P}\Big(\sum_{i=1}^{2n} X_{n,i} \leqq 0\Big) = \frac{1}{2} + \int_0^{\alpha_n} \Phi(-y/\alpha_n)\varphi(y)dy$$

and this implies

$$\Big|\mathsf{P}\Big(\sum_{i=1}^{2n} X_{n,i} \leqq 0\Big) - \Phi(0)\Big| \geqq \Phi(-1/2) \int_{\alpha_n/2}^{\alpha_n} \varphi(y)dy \geqq C\alpha_n. \qquad \square$$

2. Proofs

To simplify somewhat the notation we define for a random variable X and $t > 0$

$$D(X) := \sup\{|\mathsf{P}(X \leqq u) - \Phi(u)|;\ u \in \mathbb{R}\}$$
$$d(X,t) := \sup\{|\mathsf{P}(X \leqq u) - \Phi(u)|;\ u \geqq t\}.$$

For the proofs of the Theorem and the Corollaries we will need the following Lemma, whose proof is an easy technical exercise and therefore omitted.

LEMMA. *Let X and Y be r.v. Then*

(i) $D(Y) \leqq D(X) + (2\pi)^{-1/2}a + \mathsf{P}(|X-Y| > a) \quad \forall a > 0$
$d(Y,t) \leqq d(X,t-a) + a\varphi(t-a) + \mathsf{P}(|X-Y| > a) \quad \forall 0 < a < t.$

(ii) *For any $s > 1$ with $\mathsf{E}(|X-Y|^s) \leqq 1$ there exists a constant $C(s)$ such that*

$$D(Y) \leqq D(X) + C(s)\mathsf{E}(|X-Y|^s)^{1/(1+s)}$$

and for all $t > 0$

$$d(Y,t) \leqq d(X,t/2) + C(s)\mathsf{E}(|X-Y|^s)^{1/(1+s)}t^{-s}.$$

Furthermore we need the following Lemma 2 of [4]:

LEMMA 2. *Let X and Y be r.v., $K > 0$, $p > 1$.*

(i) *There exists a finite constant C_p such that*

$$D(X) \leqq C_p\left(D(X+Y) + \|\mathsf{E}(|Y|^p/X)\|_\infty^{1/p}\right).$$

(ii) *Let* $\|\mathsf{E}(|Y|^p/X)\|_\infty \leqq K$. *There exists a finite constant* $C_{p\,K}$ *such that for all* $x > 0$

$$d(X,x) \leqq d(X+Y,x/2) + C_{p,K}\left(D(X) + \|E(|Y|^p/X)\|_\infty^{1/p} + \|\mathsf{E}(|Y|^p/X)\|_\infty\right)$$

and

$$d(X+Y,x) \leqq d(X,x/2) + C_{p,K}\left(D(X) + \|\mathsf{E}(|Y|^p/X)\|_\infty^{1/p} + \|\mathsf{E}(|Y|^p/X)\|_\infty\right).$$

PROOF OF THE THEOREM. Let $C = C(\delta, r)$ and $t := \max\{\delta, r\}$. First we prove (i). We define a stopping time τ by

$$\tau := \sup\left\{l \in \{0,\dots,n\}; \sum_{i=1}^{l} \mathsf{E}(X_i^2/\mathcal{F}_{i-1}) \leqq 1\right\}$$

and set $Y_i := X_i I(\tau \geqq i)$ for $i = 1,\dots,n$. For $a > 0$ and $i = 1,\dots,n$ let

$$\overline{X}_i := X_i I(|X_i| \leqq a) - \mathsf{E}\,(X_i I(|X_i| \leqq a)/\mathcal{F}_{i-1}),$$
$$\overline{\overline{X}}_i := X_i I(|X_i| > a) - \mathsf{E}\,(X_i I(|X_i| > a)/\mathcal{F}_{i-1}).$$

Then we have

$$\mathsf{P}\Big(\Big|\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} Y_i\Big| > 8a\Big) \leqq$$

$$\leqq \mathsf{P}\Big(\Big|\sum_{i=\tau+2}^{n} \overline{X}_i\Big| > 2a\Big) + \mathsf{P}\Big(\Big|\sum_{i=\tau+2}^{n} \overline{\overline{X}}_i\Big| > 2a\Big) + Ca^{-2-2\delta}\sum_{i=1}^{n} \mathsf{E}(|X_i|^{2+2\delta}).$$

For the second summand on the r.h.s. we get

$$\mathsf{P}\Big(\Big|\sum_{i=\tau+2}^{n} \overline{\overline{X}}_i\Big| > 2a\Big) \leqq a^{-2}\mathsf{E}\Big(\Big|\sum_{i=\tau+2}^{n} \overline{\overline{X}}_i\Big|^2\Big) =$$

$$= a^{-2}\mathsf{E}\Big(\sum_{i=\tau+2}^{n} \overline{\overline{X}}_i^2\Big) \leqq a^{-2-2\delta} L_{n,2\delta}.$$

Now consider $\mathsf{P}\Big(\Big|\sum_{i=\tau+2}^{n} \overline{X}_i\Big| > 2a\Big)$. For $k = 1,\dots,n$ let

$$W_k := \overline{X}_k I\Big(\sum_{i=\tau+2}^{k} \mathsf{E}\left(\overline{X}_i^2/\mathcal{F}_{i-1}\right) \leqq a^2\Big) I(\tau \leqq k-2)$$

$$Z_k := \overline{X}_k I\Big(\sum_{i=\tau+2}^{k} \mathsf{E}\left(\overline{X}_i^2/\mathcal{F}_{i-1}\right) > a^2\Big) I(\tau \leqq k-2).$$

The W_k and Z_k are martingale difference sequences and with a well-known inequality of Rosenthal (see [2], Theorem 2.11) it follows

$$\mathsf{P}\Big(\Big|\sum_{i=\tau+2}^{n} \overline{X}_i\Big| > 2a\Big) \leqq$$

$$\leqq \mathsf{P}\Big(\Big|\sum_{i=\tau+2}^{n} W_i\Big| > a\Big) + \mathsf{P}\Big(\Big|\sum_{i=\tau+2}^{n} Z_i\Big| > a\Big) \leqq$$

$$\leqq a^{-2-2t}\mathsf{E}\Big(\Big|\sum_{i=\tau+2}^{n} W_i\Big|^{2+2t}\Big) + a^{-2}\mathsf{E}\Big(\Big|\sum_{i=\tau+2}^{n} Z_i\Big|^{2}\Big) \leqq$$

$$\leqq Ca^{-2-2t}\mathsf{E}\Big(\Big|\sum_{i=\tau+2}^{n} \mathsf{E}(W_i^2/\mathcal{F}_{i-1})\Big|^{1+t}\Big) + Ca^{-2-2t}\mathsf{E}\Big(\max_{\tau+2\leqq i} |W_i|^{2+2t}\Big) +$$

$$+a^{-2}\mathsf{E}\Big(\sum_{i=\tau+2}^{n} Z_i^2\Big) \leqq$$

$$\leqq Ca^{-2-2t}\mathsf{E}\Big(\Big[\sum_{k=\tau+2}^{n} \mathsf{E}(\overline{X}_k^2/\mathcal{F}_{k-1}) I\Big(\sum_{i=\tau+2}^{k} \mathsf{E}(\overline{X}_i^2/\mathcal{F}_{i-1}) \leqq a^2\Big)\Big]^{1+t}\Big) +$$

$$+Ca^{-2-2t}\mathsf{E}\Big(\max_{1\leqq k\leqq n} |\overline{X}_k|^{2+2t}\Big) +$$

$$+Ca^{-2}\mathsf{E}\Big(\sum_{k=\tau+2}^{n} \mathsf{E}(\overline{X}_k^2/\mathcal{F}_{k-1}) I\Big(\sum_{i=\tau+2}^{k} \mathsf{E}(\overline{X}_i^2/\mathcal{F}_{i-1}) > a^2\Big)\Big) =: I.$$

In the first summand $[\ldots] \leqq a^2$, and therefore $a^{-2-2t}[\ldots]^{1+t} \leqq a^{-2-2r}[\ldots]^{1+r}$. Furthermore $|\overline{X}_i| \leqq 2a$ implies

$$a^{-2-2t}\mathsf{E}\Big(\max_{1\leqq k\leqq n} |\overline{X}_k|^{2+2t}\Big) \leqq 4^{t-\delta}a^{-2-2\delta}\mathsf{E}\Big(\max_{1\leqq k\leqq n} |\overline{X}_k|^{2+2\delta}\Big).$$

So we get

$$I \leqq Ca^{-2-2r}\mathsf{E}\Big(\Big[\sum_{k=\tau+2}^{n} \mathsf{E}(\overline{X}_k^2/\mathcal{F}_{k-1}) I\Big(\sum_{i=\tau+2}^{k} \mathsf{E}(\overline{X}_i^2/\mathcal{F}_{i-1}) \leqq a^2\Big)\Big]^{1+r}\Big) +$$

$$+Ca^{-2-2\delta}L_{n,2\delta} + Ca^{-2}\mathsf{E}\Big(\sum_{k=\tau+2}^{n} \mathsf{E}(\overline{X}_k^2/\mathcal{F}_{k-1}) I\Big(\sum_{i=\tau+2}^{n} \mathsf{E}(\overline{X}_i^2/\mathcal{F}_{i-1}) > a^2\Big)\Big) \leqq$$

$$\leqq Ca^{-2-2r}\mathsf{E}\Big(\Big[\sum_{k=\tau+2}^{n}\mathsf{E}(\overline{X}_k^2/\mathcal{F}_{k-1})\Big]^{1+r}\Big)+Ca^{-2-2\delta}L_{n,2\delta}+$$

$$+Ca^{-2-2r}\mathsf{E}\Big(\Big[\sum_{k=\tau+2}^{n}\mathsf{E}(\overline{X}_k^2/\mathcal{F}_{k-1})\Big]^{1+r}\Big)\leqq$$

$$\leqq Ca^{-2-2r}\mathsf{E}\Big(\Big[\sum_{i=\tau+2}^{n}\mathsf{E}(X_i^2/\mathcal{F}_{i-1})\Big]^{1+r}\Big)+Ca^{-2-2\delta}L_{n,2\delta}\leqq$$

$$\leqq Ca^{-2-2r}\mathsf{E}\Big(\Big|\sum_{i=1}^{n}\mathsf{E}(X_i^2/\mathcal{F}_{i-1})-1\Big|^{1+r}\Big)+Ca^{-2-2\delta}L_{n,2\delta}.$$

So we have shown that

$$\mathsf{P}\Big(\Big|\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}Y_i\Big|>8a\Big)\leqq Ca^{-2-2r}N_{n,2r}+Ca^{-2-2\delta}L_{n,2\delta}.$$

Applying the Lemma with $a:=L_{n,2\delta}^{1/(3+2\delta)}+N_{n,2r}^{1/(3+2r)}$ we obtain

$$D\Big(\sum_{i=1}^{n}X_i\Big)\leqq D\Big(\sum_{i=1}^{n}Y_i\Big)+CL_{n,2\delta}^{1/(3+2\delta)}+CN_{n,2r}^{1/(3+2r)} \tag{2.1a}$$

and for $x>0$ with $a:=\Big(L_{n,2\delta}^{1/(3+2\delta)}+N_{n.2r}^{1/(3+2r)}\Big)\,x/4$

$$d\Big(\sum_{i=1}^{n}X_i,x\Big)\leqq d\Big(\sum_{i=1}^{n}Y_i,x/2\Big)+C\left(L_{n,2\delta}^{1/(3+2\delta)}+N_{n,2r}^{1/(3+2r)}\right)x^{-2-2s}. \tag{2.1b}$$

Let $0<\beta<1$ and $(Y_i)_{i>n}$ be a r.v. with $\mathsf{P}(Y_i=\pm\beta)=1/2$ and $\mathcal{F}_n, Y_{n+1}$, $Y_{n+2},\dots$ be independent. For $i>n$ let $\mathcal{F}_i:=\sigma(\mathcal{F}_n,Y_{n+1},\dots,Y_i)$,

$$\nu:=\max\Big\{l\in\mathbb{N},\sum_{i=1}^{l}\mathsf{E}(Y_i^2/\mathcal{F}_{i-1})\leqq 1\Big\}.$$

Of course $n\leqq\nu\leqq n+[\beta^{-2}]=:N-1$.

For $1\leqq i\leqq N-1$ let $Z_i:=Y_iI(\nu\geqq i)$ and

$$Z_N:=Y_N\beta^{-1}\Big(1-\sum_{i=1}^{\nu}\mathsf{E}(Y_i^2/\mathcal{F}_{i-1})\Big)^{1/2}.$$

Then we get for all $a>0$

$$\mathsf{P}\Big(\Big|\sum_{i=1}^{n}Y_i-\sum_{i=1}^{N}Z_i\Big|>a\Big)\leqq$$

$$\leqq a^{-2-2r}\mathsf{E}\Big(\Big|\sum_{i=n+1}^{N} Z_i\Big|^{2+2r} I(\tau=n)\Big) + a^{-2-2\delta}\mathsf{E}\Big(\Big|\sum_{i=n+1}^{N} Z_i\Big|^{2+2\delta} I(\tau<n)\Big) =: I.$$

Because $(Z_i I(\tau<n), \mathcal{F}_i, n+1 \leqq i \leqq N)$ and $(Z_i I(\tau=n), \mathcal{F}_i, n+1 \leqq i \leqq N)$ are mds, we obtain by a well-known inequality of Burkholder (see [2], Theorem 2.10)

$$I \leqq Ca^{-2-2r}\mathsf{E}\Big(\Big|\sum_{i=n+1}^{N} Z_i^2\Big|^{1+r} I(\tau=n)\Big) + Ca^{-2-2\delta}\mathsf{E}\Big(\Big|\sum_{i=n+1}^{N} Z_i^2\Big|^{1+\delta} I(\tau<n)\Big) \leqq$$

$$\leqq Ca^{-2-2r}N_{n,2r} + Ca^{-2-2\delta}L_{n,2\delta}.$$

In the last step we used

$$\Big|\sum_{i=n+1}^{N} Z_i^2 I(\tau=n)\Big| = \Big[1 - \sum_{i=1}^{n}\mathsf{E}(X_i^2/\mathcal{F}_{i-1})\Big] I\Big(\sum_{i=1}^{n}\mathsf{E}(X_i^2/\mathcal{F}_{i-1}) \leqq 1\Big)$$

and

$$\Big|\sum_{i=n+1}^{N} Z_i^2 I(\tau<n)\Big| \leqq \max_{1\leqq i\leqq n}\mathsf{E}(X_i^2/\mathcal{F}_{i-1}).$$

We have for all $a>0$

$$\text{(2.2)} \qquad \mathsf{P}\Big(\Big|\sum_{i=1}^{n} Y_i - \sum_{i=1}^{N} Z_i\Big| > a\Big) \leqq Ca^{-2-2r}N_{n,2r} + Ca^{-2-2\delta}L_{n,2\delta}.$$

Using the Lemma with $a := L_{n,2\delta}^{1/(3+2\delta)} + N_{n,2r}^{1/(3+2r)}$ we get

$$\text{(2.3a)} \qquad D\Big(\sum_{i=1}^{n} Y_i\Big) \leqq D\Big(\sum_{i=1}^{N} Z_i\Big) + CL_{n,2\delta}^{1/(3+2\delta)} + CN_{n,2r}^{1/(3+2r)}$$

and for $x>0$ with $a := \Big(L_{n,2\delta}^{1/(3+2\delta)} + N_{n,2r}^{1/(3+2r)}\Big)\, x/8$

$$\text{(2.3b)} \qquad d\Big(\sum_{i=1}^{n} Y_i, \frac{x}{2}\Big) \leqq d\Big(\sum_{i=1}^{N} Z_i, \frac{x}{4}\Big) + C\left(L_{n,2\delta}^{1/(3+2\delta)} + N_{n,2r}^{1/(3+2r)}\right) x^{-2-2s}.$$

The variables Z_i satisfy

$$\text{(2.4)} \qquad \sum_{i=1}^{N}\mathsf{E}(Z_i^2/\mathcal{F}_{i-1}) = 1$$

and

$$\sum_{i=1}^{N} \mathsf{E}(|Z_i|^{2+2\delta}) \leqq \sum_{i=1}^{n} \mathsf{E}(|X_i|^{2+2\delta}) + \beta^{2\delta} \sum_{i=n+1}^{N} \mathsf{E}(Z_i^2) \leqq \leqq L_{n,2\delta} + \beta^{2\delta}. \tag{2.5}$$

We choose $\beta := L_{n,2\delta}^{1/(2\delta)}$ and get with Theorem 1 of [4] for all $x \in \mathbb{R}$:

$$\left|\mathsf{P}\left(\sum_{i=1}^{N} Z_i \leqq x\right) - \Phi(x)\right| \leqq C L_{n,2\delta}^{1/(3+2\delta)} \frac{1}{1+|x|^{2+2\delta}}. \tag{2.6}$$

So we have the result in the case $x \geqq 0$. The case $x < 0$ follows from the just proved case applied to the mds $(-X_i, \mathcal{F}_i, 1 \leqq i \leqq n)$.

Now we prove part (ii) and set $a := \left\| \sum_{i=1}^{n} \mathsf{E}(X_i^2/\mathcal{F}_{i-1}) \right\|_\infty$.

Let $0 < \beta < 1$ and $(X_i)_{i>n}$ r.v. so that $\mathcal{F}_n, X_{n+1}, X_{n+2}, \ldots$ are independent and $\mathsf{P}(X_i = \pm\beta) = 1/2$ for $i > n$. For $i > n$ let $\mathcal{F}_i := \sigma(\mathcal{F}_n, X_{n+1}, \ldots, X_i)$.

$$\tau := \max\left\{ l \in \mathbb{N};\ \sum_{i=1}^{l} \mathsf{E}(X_i^2/\mathcal{F}_{i-1}) \leqq a \right\}.$$

Clearly, then $n \leqq \tau \leqq n + [a\beta^{-2}] =: N-1$. For $1 \leqq i \leqq N-1$ let $Y_i := X_i I(\tau \geqq i)$ and

$$Y_N := X_N \beta^{-1} \left(a - \sum_{i=1}^{\tau} \mathsf{E}(X_i^2/\mathcal{F}_{i-1})\right)^{1/2}.$$

Now we use the Burkholder inequality and the special construction of the $Y_i, n < i \leqq N$ and get

$$\mathsf{E}\left(\left|\sum_{i=n+1}^{N} Y_i\right|^{2+2\delta}/\mathcal{F}_n\right) \leqq C\mathsf{E}\left(\left|\sum_{i=n+1}^{N} Y_i^2\right|^{1+\delta}/\mathcal{F}_n\right) \leqq \leqq C\left|a - \sum_{i=1}^{n} \mathsf{E}(X_i^2/\mathcal{F}_{i-1})\right|^{1+\delta} \leqq C N_{n,*}^{1+\delta}. \tag{2.7}$$

Now we apply [4], Lemma 2, and get

$$D\left(\sum_{i=1}^{n} X_i\right) \leqq C D\left(\sum_{i=1}^{N} Y_i\right) + C N_{n,*}^{1/2} \tag{2.8a}$$

and for $x > 0$

$$d\left(\sum_{i=1}^{n} X_i, x\right) \leqq d\left(\sum_{i=1}^{N} Y_i, \frac{x}{2}\right) + C\left(D\left(\sum_{i=1}^{N} Y_i\right) + N_{n,*}^{1/2}\right) x^{-2-2\delta}. \tag{2.8b}$$

The variables Z_i satisfy

$$\sum_{i=1}^{N} \mathsf{E}(Y_i^2/\mathcal{F}_{i-1}) = a \tag{2.9}$$

and

$$\sum_{i=1}^{N} \mathsf{E}(|Y_i|^{2+2\delta}) \leqq \sum_{i=1}^{n} \mathsf{E}(|X_i|^{2+2\delta}) + \beta^{2\delta} \sum_{i=n+1}^{N} \mathsf{E}(Y_i^2) \leqq \leqq L_{n,2\delta} + 2\beta^{2\delta} N_{n,*}. \tag{2.10}$$

We set $\beta := L_{n,2\delta}^{1/(2\delta)}$, and obtain with Theorem 1 of [4] for all $x \in \mathbf{R}$

$$\left|\mathsf{P}\left(a^{-1/2}\sum_{i=1}^{N} Y_i \leqq a^{-1/2}x\right) - \Phi(a^{-1/2}x)\right| \leqq C L_{n,2\delta}^{1/(3+2\delta)} \frac{1}{1+|x|^{2+2\delta}}. \tag{2.11}$$

Furthermore

$$|\Phi(a^{-1/2}x) - \Phi(x)| \leqq C N_{n,*}^{1/2} |x| \exp[-x^2/3]. \tag{2.12}$$

(2.11) and (2.12) imply for all $x \in \mathbf{R}$

$$\left|\mathsf{P}\left(\sum_{i=1}^{N} Y_i \leqq x\right) - \Phi(x)\right| \leqq$$

(2.13)

$$\leqq \left|\mathsf{P}\left(a^{-1/2}\sum_{i=1}^{N} Y_i \leqq a^{-1/2}x\right) - \Phi(a^{-1/2}x)\right| + |\Phi(a^{-1/2}x) - \Phi(x)| \leqq$$

$$\leqq C\left(L_{n,2\delta}^{1/(3+2\delta)} + N_{n,*}^{1/2}\right) \frac{1}{1+|x|^{2+2\delta}},$$

and so we have shown part (ii) in the case $x \geqq 0$. The case $x < 0$ follows from the just proved applied to the mds $(-X_i, \mathcal{F}_i, 1 \leqq i \leqq n)$. □

Proof of Corollary 1. Let for $1 \leqq i \leqq n$

$$\overline{X}_i := X_i I(|X_i| \leqq \beta) - \mathsf{E}(X_i I(|X_i| \leqq \beta)/\mathcal{F}_{i-1}),$$

$$\overline{\overline{X}}_i := X_i I(|X_i| > \beta) - \mathsf{E}(X_i I(|X_i| > \beta)/\mathcal{F}_{i-1}).$$

Then

$$\mathsf{E}\left(\left|\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \overline{X}_i\right|^2\right) \leqq \mathsf{E}\left(\sum_{i=1}^{n} X_i^2 I(|X_i| > \beta)\right),$$

and therefore with the Lemma

$$D\Big(\sum_{i=1}^{n} X_i\Big) \leqq D\Big(\sum_{i=1}^{n} \overline{X}_i\Big) + CL(n,\beta)^{1/3} \tag{2.14a}$$

and for $x > 0$

$$d\Big(\sum_{i=1}^{n} X_i, x\Big) \leqq d\Big(\sum_{i=1}^{n} \overline{X}_i, \frac{x}{2}\Big) + CL(n,\beta)^{1/3}\frac{1}{1+x^2}. \tag{2.14b}$$

Using

$$\mathsf{E}\Big(\Big|\sum_{i=1}^{n} \mathsf{E}(\overline{X}_i^2/\mathcal{F}_{i-1}) - 1\Big|\Big) \leqq$$

$$\mathsf{E}\Big(\Big|\sum_{i=1}^{n} \mathsf{E}(\overline{X}_i^2/\mathcal{F}_{i-1}) - \sum_{i=1}^{n} \mathsf{E}(X_i^2/\mathcal{F}_{i-1}\Big|\Big) + \mathsf{E}\Big(\Big|\sum_{i=1}^{n} \mathsf{E}(X_i^2/\mathcal{F}_{i-1}) - 1\Big|\Big) \leqq$$

$$\leqq 2\mathsf{E}\Big(\sum_{i=1}^{n} X_i^2 I(|X_i| > \beta)\Big) + N_{n,0} = 2L(n,\beta) + N_{n,0} \quad (< 1/2),$$

we apply the Theorem with $r = 0$ and obtain

$$\Big|\mathsf{P}\Big(\sum_{i=1}^{n} \overline{X}_i \leqq x\Big) - \Phi(x)\Big| \leqq$$

$$\leqq C\left\{\Big[\sum_{i=1}^{n} \mathsf{E}(|\overline{X}_i|^{2+2\delta})\Big]^{1/(3+2\delta)} + \mathsf{E}\Big(\Big|\sum_{i=1}^{n} \mathsf{E}(\overline{X}_i^2/\mathcal{F}_{i-1}) - 1\Big|\Big)^{1/3}\right\}\frac{1}{1+x^2} \leqq$$

$$\leqq C\left\{\Big[\sum_{i=1}^{n} \mathsf{E}(|\overline{X}_i|^{2+2\delta})\Big]^{1/(3+2\delta)} + L(n,\beta)^{1/3} + N_{n,0}^{1/3}\right\}\frac{1}{1+x^2}.$$

The fact $\sum_{i=1}^{n} \mathsf{E}(|\overline{X}_i|^{2+2\delta}) \leqq 2^{1+2\delta} \sum_{i=1}^{n} \mathsf{E}(|X_i|^{2+2\delta} I(|X_i| \leqq \beta))$ implies for all $x \in \mathbb{R}$

$$\Big|\mathsf{P}\Big(\sum_{i=1}^{n} \overline{X}_i \leqq x\Big) - \Phi(x)\Big| \leqq$$

$$\leqq C\left\{\Big[\sum_{i=1}^{n} \mathsf{E}(|X_i|^{2+2\delta} I(|X_i| \leqq \beta))\Big]^{1/(3+2\delta)} + L(n,\beta)^{1/3} + N_{n,0}^{1/3}\right\}\frac{1}{1+x^2}. \quad \square$$

Proof of Corollary 2. We use Corollary 1 with $\delta = 1/2$, $\beta = 1$:

$$\left|\mathsf{P}\left(\sum_{i=1}^{n} X_i < x\right) - \Phi(x)\right| \leqq$$

$$\leqq C\left\{\left[\sum_{i=1}^{n} \mathsf{E}(|X_i|^3 I(|X_i| \leqq 1))\right]^{1/4} + L(n,1)^{1/3} + N_{n,0}^{1/3}\right\} \frac{1}{1+x^2},$$

if $N_{n,0} < 1/8$, $L(n,1) < 1/8$ and $\sum_{i=1}^{n} \mathsf{E}(|X_i|^3 I(|X_i| \leqq 1)) < 1/16$. Of course $L(n,1) \leqq W_n$. Because $|X_i| I(|X_i| \leqq 1) \leqq \int_0^1 I(|X_i| > y)dy$, we get

$$\sum_{i=1}^{n} \mathsf{E}(|X_i|^3 I(|X_i| \leqq 1)) \leqq \sum_{i=1}^{n} \mathsf{E}\left(X_i^2 \int_0^1 I(|X_i| > y)dy\right) =$$

$$= \int_0^1 \sum_{i=1}^{n} \mathsf{E}(X_i^2 I(|X_i| > y))dy = \int_0^1 L(n,y)dy = W_n \qquad \left(< \frac{1}{16}\right). \qquad \square$$

Acknowledgement. I am grateful to Prof. P. Gänßler for his support during the preparation of my thesis, which the present paper is part of.

REFERENCES

[1] Brown, B. M., Martingale central limit theorems, *Ann. Math. Statist.* **42** (1971), 59–66. *MR* **44** #7609

[2] Hall, P. and Heyde, C. C., *Martingale limit theory and its application*, Probability and Mathematical Statistics, Academic Press, New York-London, 1980. *MR* **83a**: 60001

[3] Haeusler, E., On the rate of convergence in the central limit theorem for martingales with discrete and continuous time, *Ann. Probab.* **16** (1988), 275–299. *MR* **89a**: 60060

[4] Haeusler, E. and Joos, K., A nonuniform bound on the rate of convergence in the martingale central limit theorem, *Ann. Probab.* **16** (1988), 1699–1720. *MR* **89h**: 60038

[5] Michel, R., Nonuniform central limit bounds with application to probabilities of deviations, *Ann. Probability* **4** (1976), 102–106. *MR* **52** #12047

[6] Móri, T., On the rate of convergence in the martingale central limit theorem, *Studia Sci. Math. Hungar.* **12** (1977), 413–417. *MR* **82e**: 60040

(Received July 10, 1989)

MATHEMATISCHES INSTITUT DER
UNIVERSITÄT MÜNCHEN
THERESIENSTRASSE 39
D/W-8000 MÜNCHEN 2
FEDERAL REPUBLIC OF GERMANY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 159–162

TWO COMMUTATIVITY PROBLEMS FOR RINGS

H. E. BELL[1] and A. A. KLEIN

Abstract

Polynomial identities of the form $[x^n, y] = nx^{n-1}[x, y]$ frequently occur as intermediate steps in commutativity proofs for rings. In Section 1 of this paper, we explore the commutativity implications of the weaker condition

(†) for each element x of the ring R, there exists an integer $n = n(x) > 1$ such that $[x^n, y] = nx^{n-1}[x, y]$ for all $y \in R$.

Our motivation for studying (†) was an attempt to prove a theorem extending a recent result of Abu-Khuzam and Yaqub [2, Theorem 2], and we present this theorem in Section 2.

Throughout the paper, R will denote a ring with center Z and commutator ideal $C(R)$. As usual, $[x, y]$ will denote the commutator $xy - yx$.

1. Rings satisfying (†)

THEOREM 1. *If R is any ring satisfying (†), then $C(R)$ is nil.*

PROOF. We need only establish commutativity in the case of R with no nonzero nil ideals; and since such a ring is a subdirect product of prime rings with no nonzero nil ideals, we assume henceforth that R is prime with no nonzero nil ideals. We shall show that R is radical over Z, in which case commutativity follows by an old theorem of Herstein [4].

Fix $x \in R$, and let $n = n(x)$. If $\operatorname{char} R = p$ and $p \mid n$, it is immediate from (†) that $x^n \in Z$; therefore, we assume that either $p \nmid n$ or $\operatorname{char} R = 0$. Replacing y by yz in (†), we get

$$[x^n, y]z + y[x^n, z] = nx^{n-1}[x, y]z + nx^{n-1}y[x, z];$$

and using (†), we get

$$y[x^n, z] = nx^{n-1}y[x, z].$$

1980 *Mathematics Subject Classification* (1985 Revision). Primary 16A70; Secondary 16A12.

Key words and phrases. Commutativity, semiprime rings

[1]Supported by the Natural Sciences and Engineering Research Council of Canada Grant No. A 3961.

Replacing the left side of this equality by $nyx^{n-1}[x,z]$ now yields $n[x^{n-1},y][x,z]=0$, hence

$$[x^{n-1},y][x,z]=0 \quad \text{for all} \quad y,z\in R. \tag{1}$$

Substituting zw for z in (1) gives

$$[x^{n-1},y]R[x,w]=\{0\},$$

and primeness of R implies that either $x\in Z$ or $[x^{n-1},y]=0$ for all $y\in R$. Thus $x^{n-1}\in Z$, and R is radical over Z as claimed.

The existence of prime nil rings shows that (†) does not imply commutativity, even in prime rings; however, under additional hypotheses we can indeed prove commutativity.

For each $x\in R$, define N_x to be the set of all integers $n\geqq 2$ for which $[x^n,y]=nx^{n-1}[x,y]$ for all $y\in R$. If R satisfies (†), it is easy to show that N_x is infinite for each $x\in R$.

THEOREM 2. *Let R be a 2-torsion-free semiprime ring with 1. If each N_x contains three consecutive integers (depending on x), then R is commutative.*

PROOF. Suppose $u\in R$ satisfies $u^2=0\neq u$, and let $n, n+1$ and $n+2$ be elements of N_{1+u}. The condition that $[(1+u)^n,y]=n(1+u)^{n-1}[1+u,y]$ reduces at once to

$$n(n-1)u[u,y]=0 \quad \text{for all} \quad y\in R; \tag{2}$$

and replacing n by $n+1$ gives

$$(n+1)nu[u,y]=0 \quad \text{for all} \quad y\in R. \tag{3}$$

Subtracting (2) from (3) yields $2nu[u,y]=0$; and since we can also obtain $2(n+1)u[u,y]=0$ by applying the same arguments for $n+1$ and $n+2$, we conclude that $2u[u,y]=0$ for all $y\in R$. Thus, $u[u,y]=0=uyu$ for all $y\in R$, so that $u=0$ by semiprimeness of R. Hence R has no nonzero nilpotent elements, and commutativity of R follows from Theorem 1.

Letting R satisfy (†), for each $x\in R$ define d_x to be the g.c.d. of the set $\{n(n-1)\mid n\in N_x\}$. Then $2\mid d_x$ for each $x\in R$. Moreover, a careful analysis of the proof of Theorem 2 reveals that the three-consecutive-n hypothesis can be replaced by the hypothesis that R is d_x-torsion-free for each $x\in R$. The latter hypothesis can be satisfied in a variety of ways, each yielding a variant of Theorem 2. For example, since $(n(n-1),(n+2)(n+1))$ is either 2 or 6 for each $n\geqq 2$, we obtain

THEOREM 3. *Let R be a 6-torsion-free semiprime ring with 1. If each N_x contains a pair of integers differing by 2, then R is commutative.*

2. An application

Abu-Khuzam has proved the following pretty theorem, which was generalized somewhat in [2]:

THEOREM A–K ([1]). *If R is a semiprime ring such that for each $x \in R$ there exists $n = n(x) > 1$ for which $(xy)^n = x^n y^n$ for all $y \in R$, then R is commutative.*

It is our purpose to prove an extension of this result — specifically,

THEOREM 4. *Let R be a semiprime ring with the property that*

(††) *for each $x \in R$, there exists an integer $n = n(x) > 1$ such that $(xy)^n - x^n y^n \in Z$ for all $y \in R$.*

Then R is commutative.

PROOF. Since R is a subdirect product of prime rings, we assume from the beginning that R is prime, in which case R is a domain by [2, Lemma 3]. By (††) and Theorem A–K, it is clear that $Z \neq \{0\}$; and we can localize at $Z \setminus \{0\}$, obtaining a domain R^* with 1, in which R is embedded. It is readily seen that R^* inherits (††); moreover, as we now show, R^* is a division ring.

We need only show that for each $x \in R \setminus \{0\}$, there exists $u \in R$ such that $xu \in Z \setminus \{0\}$; and this will be the case if for each $x \in R \setminus \{0\}$, there exists $w \in R$ such that $(xwy)^{n(xw)} - (xw)^{n(xw)} y^{n(xw)} \neq 0$ for some $y \in R$. The alternative is that there exists $x \in R \setminus \{0\}$ such that for each $v \in xR$, there exists $m = m(v) > 1$ such that $(vy)^m = v^m y^m$ for all $y \in R$, and in particular for all $y \in xR$; and xR is therefore commutative by Theorem A–K. But a domain R with a nonzero commutative right ideal is itself commutative, which implies that R^* is a field.

From now on, we assume that R is a division ring. Fixing $x \in R \setminus \{0\}$ and taking $n = n(x)$, we see from (††) that

$$xyx^n y^n = x^n y^n xy \quad \text{for all} \quad y \in R; \tag{4}$$

thus, R satisfies a generalized polynomial identity (GPI). Now Amitsur has shown that a division ring R satisfying a GPI is finite-dimensional over Z [3, Theorem 13]; consequently, if Z is finite, R is finite, and thus commutative by Wedderburn's theorem.

It remains only to treat the case of a division ring with Z infinite. We return to (4), which we rewrite as

$$x[x^n, y]y^n + x^n[y^n, x]y = 0 \quad \text{for all} \quad y \in R. \tag{5}$$

Substitute $y + 1$ for y in (5), obtaining

$$x[x^n, y](y+1)^n + x^n[(y+1)^n, x](y+1) = 0 \quad \text{for all} \quad y \in R. \tag{6}$$

Expanding (6) and collecting terms of the same y-degree, we obtain the equation $\sum_{i=1}^{n+1} w_i(y) = 0$, where $w_i(y)$ denotes the sum of the terms of y-degree i.

Replacing y by λy for $n+1$ different $\lambda \in Z \setminus \{0\}$, and using a standard Vandermonde argument, we see that $w_i(y) = 0$ for all $y \in R$ and all $i = 1, 2, \ldots, n+1$. But examination of (6) shows that $w_1(y) = x[x^n, y] + nx^n[y, x]$; and by equating to zero and cancelling an x, we get $[x^n, y] - nx^{n-1}[x, y] = 0$. Thus, R satisfies (†) and is commutative by Theorem 1.

REFERENCES

[1] ABU-KHUZAM, H., A commutativity theorem for semiprime rings, *Bull. Austral. Math. Soc.* **27** (1983), 221–224. *MR* **84i**: 16037

[2] ABU-KHUZAM, H. and YAQUB, A., Commutativity of certain semiprime rings, *Studia Sci. Math. Hungar.* **24** (1989), 33–36. *MR* **90b**: 16041

[3] AMITSUR, S. A., Generalized polynomial identities and pivotal monomials, *Trans. Amer. Math. Soc.* **114** (1965), 210–226. *MR* **30** #3117

[4] HERSTEIN, I. N., A theorem on rings, *Canadian J. Math.* **5** (1953), 238–241. *MR* **14**–719

(Received July 12, 1989)

MATHEMATICS DEPARTMENT
BROCK UNIVERSITY
ST. CATHARINES, ONTARIO
L2S 3A1
CANADA

SCHOOL OF MATHEMATICAL SCIENCES
TEL AVIV UNIVERSITY
RAMAT AVIV
IL–69978 TEL-AVIV
ISRAEL

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 163–172

COVERING OF A TRIANGLE BY HOMOTHETIC TRIANGLES

É. VÁSÁRHELYI

Let B be a convex domain in the Euclidean plane and $\{B_i\}_{i=1}^N$ a finite set of homothetic copies of B. L. Fejes Tóth proposed the following problem: how large must be the sum of the areas of the B_i, so that B can be covered by translates of the B_i?

We shall deal here with a related problem. (For earlier results, see references [1], [2], [3], [4].)

Consider a triangle T, the area of which will also be denoted by T, in the Euclidean plane and another triangle T^φ, which is obtained from T by a rotation through a given angle φ. The problem we are interested in is to determine $f_\varphi(T)$, the minimal number with the following property: if $\{T_i^\varphi\}_{i=1}^N$ is an arbitrary finite set of homothetic copies of T^φ with total area at least $Tf_\varphi(T)$, then T can be covered by translates of T_i^φ $(i = 1, \dots, N)$.

We note that for the case $\varphi = 180°$ it was proved in [4] that $f_{180°}(T) = 4$, which implies the inequality

$$(*) \qquad \max_\varphi f_\varphi(T) \geqq 4.$$

A. Bezdek and K. Bezdek [1] conjectured that equality holds if and only if T is a regular triangle.

In the original problem one can neglect the metrical difference of affine equivalent domains but in this problem we try to characterize the regular triangle among all triangles by the equality in $(*)$.

First we show that

$$\max_\varphi f_\varphi(T) > 4,$$

for any non-regular triangle.

In order to prove this, let us note that, for a triangle $T = ABC$ with at least two different angles (for example $\beta \geqq \alpha > \gamma$), there is an angle $\varphi = \alpha$ for which (in case $N = 1$) the triangle T cannot be covered by homothetic copies of T^α with total area $4T$, that is $f_\alpha(T) > 4$.

The second part of the conjecture is connected with regular triangles. Since a regular triangle Δ has a rotational symmetry of order 3, it is sufficient

1980 *Mathematics Subject Classification.* Primary 52A45.

Key words and phrases. Covering, triangle, homothetical covering.

Akadémiai Kiadó, Budapest

to examine only the values $\varphi \in [0°; 60°]$. For this, we have only the following partial results (a), (b), (c).

A. Bezdek and Z. Füredi [3] proved that, for any triangle T

$$f_{0°}(T) = 2.$$

In the case of regular triangles Δ we consider the triangles Δ_i inscribed in Δ_i^{φ}, where Δ_i is a homothetic copy of Δ.

If $\varphi \in [0°; 15°]$, then $\Delta_i \geqq \frac{1}{2}\Delta_i^{\varphi}$, and we have

$$\sum \Delta_i \geqq \frac{1}{2} \sum \Delta_i^{\varphi},$$

whence

(a) $$f_{\varphi}(\Delta) \leqq 4, \quad \text{for} \quad \varphi \in [0°; 15°].$$

The case $\varphi = 60°$ corresponds to the covering by triangles in the "opposite" position. It was proved in [4] that $f_{60°}(T) = 4$, thus

(b) $$f_{60°}(\Delta) = 4.$$

Now we shall show that, for regular triangles,

(c) $$f_{30°}(\Delta) \leqq 4.$$

THEOREM. *Let Δ be a regular triangle in the Euclidean plane and Δ' another one obtained by rotation of Δ through the angle 30°. If $\{\Delta_i'\}_{i=1}^{N}$ is a set of homothetic copies of Δ' with total area at least 4Δ, then Δ can be covered by translates of Δ_i' $(i = 1, \ldots, N)$.*

In order to prove the theorem we need two auxiliary results.

LEMMA 1. *Consider a regular triangle Δ $(= A_0B_0C_0)$ with side length 1. Let $R_1, \ldots, R_N$ be rectangles with side length c_i, $c_i\frac{\sqrt{3}}{2}$ $(i = 1, \ldots, N;\ c_1 \geqq c_2 \geqq \ldots \geqq c_N)$, whose side of length c_i is perpendicular to B_0C_0. Let T $(= BCAA_0)$ be a trapezoid determined by*

$$A_0A \| B_0C_0,\ A_0A = c\frac{\sqrt{3}}{6}; \quad B \in A_0B_0, \quad A_0B = 1 + \frac{2\sqrt{3}}{3}c;$$

$BC \| B_0C_0$, $AC \| A_0C_0$ with $c \geqq c_1$ (Fig. 1). If the total area of the rectangles is not smaller than the area of the trapezoid T, i.e.

(1) $$\sum_{i=1}^{N} R_i \geqq T = \left[\left(1 + \frac{5}{6}c\sqrt{3}\right)^2 - \left(\frac{1}{6}c\sqrt{3}\right)^2\right]\Delta,$$

then the triangle $A_0B_0C_0$ can be covered by translates of the rectangles R_i.

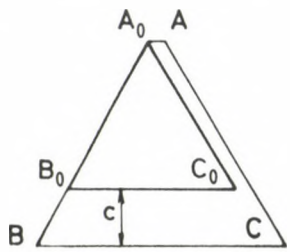


Fig. 1

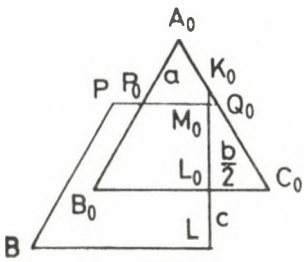


Fig. 2

LEMMA 2. *Consider a regular triangle Δ $(= A_0B_0C_0)$ with side length* 1 *and the points P_0, Q_0, K_0, L_0, M_0, with*

$$P_0 \in A_0B_0, \quad Q_0 \in A_0C_0, \quad A_0P_0 = A_0Q_0 = a \geqq 0, \quad K_0 \in A_0C_0,$$
$$L_0 \in B_0C_0, \quad K_0C_0 = b, \quad K_0L_0 \perp B_0C_0, \quad M_0 = K_0L_0 \cap P_0Q_0.$$

Let $R_1, \dots, R_N$ be rectangles with side length c_i, $c_i\frac{\sqrt{3}}{2}$ $(i = 1, \dots, N;\ c_1 \geqq c_2 \geqq \dots \geqq c_N)$, whose side c_i is perpendicular to B_0C_0. Let $\tilde{T}$ $(= PBLM_0)$ be a trapezoid determined by $c \geqq c_1$, $M_0L = M_0L_0 + c$, $M_0P = M_0P_0 + c\frac{\sqrt{3}}{3}$, $PB \| A_0B_0$, and $LB \| B_0C_0$ (Fig. 2). If the total area of the rectangles is not smaller than the area of the trapezoid $\tilde{T}$, i.e.

$$(2) \qquad \sum_{i=1}^{N} R_i \geqq \tilde{T} = \left(2 - 2a + 4c\frac{\sqrt{3}}{3}\right)\left(\frac{3}{4} - \frac{b}{2} + \frac{a}{4} + c\frac{\sqrt{3}}{2}\right)\Delta,$$

then the trapezoid $L_0M_0P_0B_0$ can be covered by translates of $R_1, R_2, \dots, R_n$.

PROOF OF THE LEMMAS. Since the proofs are similar, we sketch the proof of Lemma 1 only.

We define classes of rectangles. Let $\mathcal{R}_1 = \{R_i \mid i = 1, \dots, j_1\}$ where j_1 is determined by

$$\frac{\sqrt{3}}{2}\sum_{i=1}^{j_1} c_i \geqq B_0C_0 > \frac{\sqrt{3}}{2}\sum_{i=1}^{j_1-1} c_i.$$

The straight line parallel to B_0C_0 at distance c_{j_1} intersects Δ in B_1 and C_1. Let $\mathcal{R}_2$ be the second class of rectangles, where

$$\mathcal{R}_2 = \{R_i \mid i = j_1 + 1, \dots, j_2\}, \quad \frac{\sqrt{3}}{2}\sum_{i=j_1+1}^{j_2} c_i \geqq B_1C_1 > \frac{\sqrt{3}}{2}\sum_{i=j_1+1}^{j_2-1} c_i.$$

The straight line parallel to B_1C_1 at distance c_{j_2} intersects Δ in B_2 and C_2. Repeat the process given above for B_2C_2 starting with the $(j_2 + 1)$-th rectangle. We obtain the trapezoids $B_{s-1}C_{s-1}C_sB_s$ and the classes of rectangles

$$\mathcal{R}_s = \{R_i \mid i = j_{s-1} + 1, \dots, j_s\},$$

where the index j_s satisfies

$$\frac{\sqrt{3}}{2}\sum_{i=j_s+1}^{j_s} c_i \geqq B_{s-1}C_{s-1} > \frac{\sqrt{3}}{2}\sum_{i=j_{s-1}+1}^{j_s-1} c_i$$

with $j_0 = 0$. This process may be continued until an index n. We have to stop if either

$$A_0 B_{n-1} \leqq \frac{2\sqrt{3}}{3} c_{j_n} \tag{3}$$

or

$$\frac{\sqrt{3}}{2}\sum_{i=j_n+1}^{N} c_i < B_n C_n. \tag{4}$$

(In the first case the straight line parallel to $B_{n-1}C_{n-1}$ at distance c_{j_n} has at most the point A_0 common with Δ. In the second case we have not enough rectangles for the new class $\mathcal{R}_{n+1}$.)

We suppose that the process has to stop because of (4). We put the rectangles one after the other in monotone decreasing sequence according to Figure 3. One can see easily that the area of the trapezoid $BCAA_0$ gives an upper bound of the total area of $\{R_i\}_{i=1}^N$, and we have $T > \sum_{i=1}^N R_i$ contrary to the hypothesis of Lemma 1. It follows that our process may be continued until (3). The trapezoids $B_0C_0C_1B_1, B_1C_1C_2B_2, \ldots, B_{n-2}C_{n-2}C_{n-1}B_{n-1}$ and the triangle $B_{n-1}C_{n-1}A_0$ can be covered by the classes $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_{n-1}$ and $\mathcal{R}_n$, respectively. □

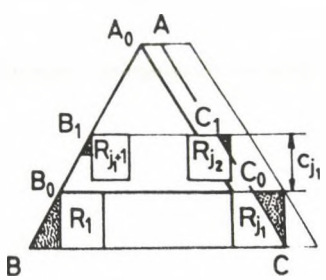


Fig. 3

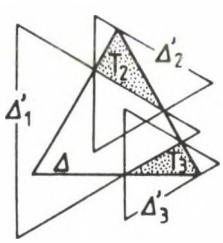


Fig. 4

PROOF OF THE THEOREM. Let Δ be a regular triangle of side 1 and let a_i denote the side of Δ_i'. In this case the condition of the theorem is equivalent to the inequality

$$\sum_{i=1}^{N} a_i^2 \geqq 4. \tag{5}$$

The proof is by induction on N. We may assume, without loss of generality, that $a_1 \geqq a_2 \ldots \geqq a_N$. We note that the triangle Δ can be covered by a homothetic copy of Δ' with side length at least $\sqrt{3}$, and a right triangle at a vertex of Δ with hypothenuse u can be covered by a homothetic copy of Δ' with side length at least $\frac{5}{6}u\sqrt{3}$.

First we prove our statement for $N = 1, 2, 3$. In cases $N = 1$, $N = 2$ we put $a_2 = a_3 = 0$ and $a_3 = 0$, respectively. Translate the triangles Δ_3' and Δ_2' so that they cover the triangles T_3 and T_2 with hypothenuse $0,4a_3\sqrt{3}$ and $0,4a_2\sqrt{3}$ respectively (Fig. 4). If $a_1 + 0,4a_2 + 0,4a_3 \geqq \sqrt{3}$, then one can easily see that Δ_1' can cover the remaining part of Δ.

We consider a_1, a_2 and a_3 as coordinates of a point in a rectangular system of coordinates. The tetrahedron with vertices

$$P_1(0;0;0),\ P_2(\sqrt{3};0;0),\ P_3\left(\frac{5}{7}\sqrt{3};\frac{5}{7}\sqrt{3};0\right),\ P_2\left(\frac{5}{9}\sqrt{3};\frac{5}{9}\sqrt{3};\frac{5}{9}\sqrt{3}\right)$$

contains all of the points $P(a_1;a_2;a_3)$, whose coordinates satisfy the conditions

$$a_1 \geqq a_2 \geqq a_3 \quad \text{and} \quad \sqrt{3} > a_1 + 0,4a_2 + 0,4a_3.$$

On the other hand, the tetrahedron $P_1P_2P_3P_4$ is contained in the interior of the ball $a_1^2 + a_2^2 + a_3^2 = 4$. Thus, if $a_1^2 + a_2^2 + a_3^2 \geqq 4$, then the point $P(a_1;a_2;a_3)$ is outside the tetrahedron, and Δ_1', Δ_2' and Δ_3' can cover Δ.

Now we suppose that our statement is true for $N < m$ and prove it for $N = m$. We assume that $\sqrt{3} > a_1 \geqq a_2 \ldots \geqq a_N$, while in the case $a_1 \geqq \sqrt{3}$, Δ can be covered by Δ_1'.

Introducing the intervals:

$$I_1 = \left[0;\frac{\sqrt{3}}{3}\right), \qquad I_2 = \left[\frac{\sqrt{3}}{3};\frac{\sqrt{3}}{2}\right),$$
$$I_3 = \left[\frac{\sqrt{3}}{2};\frac{2\sqrt{3}}{3}\right), \qquad I_4 = \left[\frac{2\sqrt{3}}{3};\sqrt{3}\right)$$

the case $a_i \in I_{j_i}$ $(i = 1,2,3)$ will be designated by the three-place number $j_1j_2j_3$. That is, we have to show in 20 cases that Δ can be covered by $\{\Delta_i'\}_{i=1}^N$, but some of them can be proved together.

Case 1. 444, 443, 442, 433, 432, 333.

If $a_1 = \frac{2\sqrt{3}}{3}$, $a_2 = \frac{\sqrt{3}}{2}$ and $a_3 = \frac{\sqrt{3}}{3}$ then the triangles Δ_1', Δ_2' and Δ_3' cover Δ (Fig. 5), and this guarantees the covering for 432, 433, 442, 443 and 444. The same is true for 333 (Fig. 6).

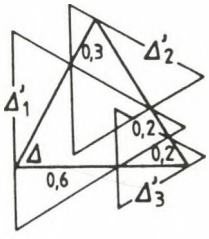


Fig. 5

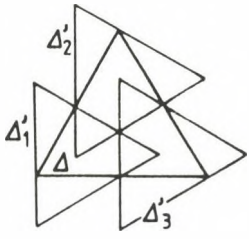


Fig. 6

CASE 2. 411.

We cover by Δ_1' a hexagon so that there are three congruent rightangled triangles at the vertices of Δ, whose points are not covered. We complete these triangles by reflexion to regular triangles with side length

$$b_1 = \frac{\sqrt{3} - a_1}{3},$$

which are homothetic to Δ. It is possible to divide the triangles $\Delta_2', \Delta_3', \ldots, \Delta_N'$ into three classes so that the total area of each class is at least $\frac{\Delta}{3}(4 - a_1^2 - 2a_2^2)$.

A simple calculation shows, for the case 411 that

$$\frac{1}{3}(4 - a_1^2 - 2a_2^2) \geqq 4b_1^2.$$

Thus, by the inductive hypothesis, each of the new triangles with side length b_1 can be covered by the triangles of a class.

CASE 3. 441, 431, 421, 331, 321.

We consider a_1 and a_2 as coordinates of a point in a rectangular system of coordinates. The trapezoid

$$P_1\left(\frac{\sqrt{3}}{2}; \frac{\sqrt{3}}{3}\right) P_2\left(\sqrt{3}; \frac{\sqrt{3}}{3}\right) P_3(\sqrt{3}; \sqrt{3}) P_4\left(\frac{\sqrt{3}}{2}; \frac{\sqrt{3}}{2}\right)$$

contains all of the points $P(a_1; a_2)$, whose coordinates are determined by the condition of this case. If $a_1 + 0,4a_2 \geqq \sqrt{3}$, then Δ can be covered by Δ_1' and Δ_2'; consequently, we may assume that $P \in P_1X_1X_2P_4$, where $X_1\left(\frac{13}{15}\sqrt{3}; \frac{\sqrt{3}}{3}\right)$, $X_2\left(\frac{5}{7}\sqrt{3}; \frac{5}{7}\sqrt{3}\right)$. Now we translate Δ_1' so that a rightangled triangle with hypothenuse $0,4a_2\sqrt{3}$ and two congruent rightangled triangle with hypothenuse

$$b_2 = \frac{3}{2} - a_1\frac{\sqrt{3}}{2} - a_2\frac{\sqrt{3}}{5}$$

remain uncovered. The first one can be covered by Δ_2'. In order to cover the remaining two, we divide the triangles $\Delta_3', \ldots, \Delta_N'$ again into classes, but

now we need just two classes. The total area of a class is at least $\frac{\Delta}{2}(4 - a_1^2 - a_2^2 - a_3^2)$. Thus, we have to show that

$$\frac{1}{2}(4 - a_1^2 - a_2^2 - a_3^2) \geqq 4b_2^2.$$

Since $a_3 < \frac{\sqrt{3}}{3}$ it is sufficient to prove that

$$7a_1^2 + \frac{49}{25}a_2^2 + \frac{24}{5}a_1a_2 - 12a_1\sqrt{3} - \frac{24}{5}a_2\sqrt{3} + \frac{43}{3} \leqq 0. \tag{6}$$

Note that (6) defines an elliptical domain containing the points P_1, X_1, X_2, P_4 in its interior; therefore, (6) holds for any $P \in P_1X_1X_2P_4$. This implies that both of the non-covered triangles can be covered by a class.

CASE 4. 422 and 322.

We cover by Δ_3' and Δ_2' a triangle with hypothenuse $0,4$ and $0,4a_2\sqrt{3}$ respectively, and we put Δ_1' in the obvious way to them. Let $b_3 = 2,6 - a_1\sqrt{3} - 0,4a_2\sqrt{3}$. If $b_3 \leqq 0$ then Δ is covered by Δ_1', Δ_2' and Δ_3'. In the opposite case there is a rightangled triangle with hypothenuse b_3, which is not covered. We cover instead of this a bigger one, a regular triangle $\bar{\Delta}$ with side length b_3, which is homothetic to Δ. From the condition of the theorem we obtain

$$\sum_{i=4}^{N} a_i^2 \geqq 4 - a_1^2 - a_2^2 - a_3^2 \geqq 4 - a_1^2 - 2a_2^2.$$

Now we have the condition of the induction in the following form

$$4 - a_1^2 - 2a_2^2 \geqq 4b_3^2,$$

which is equivalent to the inequality

$$13a_1^2 + 3,92a_2^2 + 9,6a_1a_2 - 20,8a_1\sqrt{3} - 8,32a_2\sqrt{3} + 23,04 \leqq 0. \tag{7}$$

Formula (7) defines an elliptical domain containing the convex hull of the points

$$X_1\left(\frac{\sqrt{3}}{2}; \frac{\sqrt{3}}{2}\right),\ X_2\left(\frac{\sqrt{3}}{2}; \frac{\sqrt{3}}{3}\right),\ X_3\left(\frac{2,2\sqrt{3}}{3}; \frac{\sqrt{3}}{3}\right),\ X_4\left(\frac{2\sqrt{3}}{3}; \frac{\sqrt{3}}{2}\right)$$

together with the points $P(a_1; a_2)$ whose coordinates satisfy the conditions of cases 422 and 322 with $b_3 \geqq 0$. In consequence of this the triangles $\Delta_4', \ldots, \Delta_N'$ cover $\bar{\Delta}$, and so Δ can be covered by $\Delta_1', \ldots, \Delta_N'$.

CASE 5. 332.

We cover a trapezoid in Δ by Δ_1' and Δ_2', and we have to show that the remaining regular triangle with side length $b_4 = \frac{4}{3}\left(1 - a_1\frac{\sqrt{3}}{3}\right)$ can be covered by $\Delta_3', \ldots, \Delta_N'$.

From the condition of the theorem we have

$$\sum_{i=3}^{N} a_i^2 \geqq 4 - a_1^2 - a_2^2 \geqq 4 - 2a_1^2.$$

Since $a_1, a_2 \in I_3$, the inequality $4 - 2a_1^2 \geqq 4b_4^2$ holds, which completes the proof of this case.

CASE 6. 222.

Now we have $a_i \in I_2$, $i = 1, 2, 3$. The proof is similar to those given above; we just use three of the triangles in order to cover the trapezoid and obtain a non-covered regular triangle with side length $b_5 = 1 - a_1 \frac{\sqrt{3}}{4}$. One can easily verify the inequalities

$$\sum_{i=4}^{N} a_i^2 \geqq 4 - a_1^2 - a_2^2 - a_3^2 \geqq 4 - 3a_1^2 \geqq 4b_5^2.$$

The proof of the other cases 311, 221, 211 and 111 depends upon our lemmas. In order to apply the lemmas we consider the rectangles R_i with side length $a_i \frac{\sqrt{3}}{4}$ and $\frac{1}{2}a_i$, which are inscribed in Δ_i' and have area $\frac{1}{2}\Delta_i'$ $(i = 1, \ldots, N)$.

CASE 7. 111. We make use of Lemma 1 with the values $c = \frac{\sqrt{3}}{6}$, $c_i = \frac{a_i}{2}$. The hypothesis of our theorem implies that

$$\frac{1}{2}\sum_{i=1}^{N} \Delta_i' = \sum_{i=1}^{N} R_i \geqq \left[\left(1 + \frac{5}{12}\right)^2 - \left(\frac{1}{12}\right)^2\right]\Delta = 2\Delta,$$

and thus the condition of Lemma 1 holds. The inscribed rectangles cover Δ.

CASE 8. 211, 311.

We translate the triangles Δ_1' and Δ_2' so that (corresponding to Figure 7) they cover at the vertex C_0 a rightangled triangle $K_0L_0C_0$ with hypothenuse $a_1\frac{\sqrt{3}}{2}$ and at the vertex A_0 a regular triangle $A_0P_0Q_0$ with side length $a_2\frac{\sqrt{3}}{3}$ (which is homothetic to Δ), respectively. The non-covered points of Δ are contained in the trapezoid $L_0M_0P_0B_0$. In order to complete the covering of Δ we are going to apply Lemma 2.

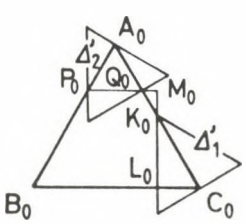


Fig. 7

Let $a = a_2\frac{\sqrt{3}}{3}$, $b = a_1\frac{\sqrt{3}}{2}$, $c_i = \frac{a_i}{2}$ $(i = 3, 4, \ldots, N)$, $c = \frac{a_2}{2}$. From the condition of the theorem we obtain

$$\sum_{i=3}^{N} R_i \geqq \frac{1}{2}(4 - a_1^2 - a_2^2)\Delta, \tag{8}$$

and we shall show that

$$\frac{1}{2}(4 - a_1^2 - a_2^2)\Delta \geqq \left(2 - 2a_2\frac{\sqrt{3}}{3} + 4a_2\frac{\sqrt{3}}{6}\right)\left(\frac{3}{4} - a_1\frac{\sqrt{3}}{4} + a_2\frac{\sqrt{3}}{12} + a_2\frac{\sqrt{3}}{4}\right)\Delta,$$

which is equivalent to

$$a_1^2 + a_2^2 - a_1\sqrt{3} + \frac{4}{3}a_2\sqrt{3} - 1 \leqq 0. \tag{9}$$

The circle in (9) contains the quadrangle

$$X_1\left(\frac{\sqrt{3}}{3}; 0\right) X_2\left(\frac{2\sqrt{3}}{3}; 0\right) X_3\left(\frac{2\sqrt{3}}{3}; \frac{\sqrt{3}}{3}\right) X_4\left(\frac{\sqrt{3}}{3}; \frac{\sqrt{3}}{3}\right).$$

(X_3, X_4 are on the boundary and X_1, X_2 are inside of it.) Therefore the coordinates of any point $P(a_1; a_2)$ considered in this case satisfy (9). From (8) and (9) we obtain that the condition of Lemma 2

$$\sum_{i=3}^{N} R_i \geqq \left(2 - 2a_2\frac{\sqrt{3}}{3} + 4a_2\frac{\sqrt{3}}{6}\right)\left(\frac{3}{4} - a_1\frac{\sqrt{3}}{4} + a_2\frac{\sqrt{3}}{12} + a_2\frac{\sqrt{3}}{4}\right)\Delta$$

holds, and the inscribed rectangles cover the remaining part of Δ.

CASE 9. 221.

The proof of Case 9 is similar to that of Case 8, but now we have to choose $c = \frac{\sqrt{3}}{6} \geqq \frac{1}{2}a_3$. The new forms of our inequalities are

$$\sum_{i=3}^{N} R_i \geqq \frac{1}{2}(4 - a_1^2 - a_2^2)\Delta \geqq \left(2 - 2a_2\frac{\sqrt{3}}{3} + \frac{2}{3}\right)\left(\frac{3}{4} - a_1\frac{\sqrt{3}}{4} + a_2\frac{\sqrt{3}}{12} + \frac{1}{4}\right)\Delta,$$

and

$$9a_1^2 + 6a_2^2 + 9a_1a_2 - 12a_1\sqrt{3} - 8a_2\sqrt{3} + 12 \leqq 0. \tag{10}$$

The points, for which (10) holds form an elliptical domain. It can easily be verified by calculation that $X_1\left(\frac{\sqrt{3}}{3}; \frac{\sqrt{3}}{3}\right)$ and $X_2\left(\frac{\sqrt{3}}{2}; \frac{\sqrt{3}}{3}\right)$ are its boundary points, and $X_3\left(\frac{\sqrt{3}}{2}; \frac{\sqrt{3}}{2}\right)$ is inside. Since the triangle $X_1X_2X_3$ contains all

of the points $P(a_1, a_2)$ considered in this case, (10) holds and this completes the proof of the theorem. □

REFERENCES

[1] BEZDEK, A. and BEZDEK, K., Eine hinreichende Bedingung für die Überdeckung des Einheitswürfels durch homothetische Exemplare im n-dimensionalen euklidischen Raum, *Beiträge Algebra Geom.* **17** (1984), 5–21. *MR* **85h**: 52017

[2] BEZDEK, A., Ausfüllung und Überdeckung der Ebene durch Kreise, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **28** (1985), 173–177. *MR* **87i**: 52023

[3] BEZDEK, A. and FÜREDI, Z., Covering a triangle with triangles (to appear).

[4] VÁSÁRHELYI, É., Über eine Überdeckung mit homothetischen Dreiecken, *Beiträge Algebra Geom.* **17** (1984), 61–70. *MR* **85h**: 52020

(Received August 1, 1989)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
GEOMETRIA TANSZÉK
RÁKÓCZI ÚT 5
H–1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 173–178

LEARNING WITH FINITE MEMORY

J. KOMLÓS,* L. REJTŐ and G. TUSNÁDY

Abstract

Let the input of a finite automaton be a sequence of independent symmetric ± 1 variables. The automaton has no output but a total payoff is given which is the modulus of the conditional expectation of the sum of input elements on the final state. Let $C(n, m)$ denote the maximal expected payoff for n inputs and m states. It is proven here that $C(n, 2)$ is bounded, while $C(n, 3) \geqq \kappa \log n$ with some positive constant κ.

1. Introduction

Let Λ be an arbitrary $N \times N$ matrix with non-negative elements. A particle is wandering on N possible states in the following way. Being in state i, $1 \leqq i \leqq N$, let $T_1, T_2, \ldots, T_N$ be independent random variables with exponential distributions with parameters $\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{iN}$, where λ_{ij}-s are the entries of Λ. If T_j is the minimum of $T_1, T_2, \ldots, T_N$, then the particle moves to state j in time T_j.

If the indices of positive elements of Λ define a strongly connected directed graph then the process defined above is ergodic. Thus Λ determines a distribution, the stationary distribution of the random walk. In the case

$$\lambda_{ij} = a_{ij} \exp(E_i - E_j), \quad 1 \leqq i, j \leqq N,$$

where $a_{ji} = a_{ij}$ is the adjacency of an undirected graph and the "energy level" $E_1, E_2, \ldots, E_N$ are arbitrary real numbers, the stationary distribution is

$$p_i = \kappa \exp(-2E_i), \quad 1 \leqq i \leqq N,$$

where κ is a norming factor. (The infinitesimal transport on edge i, j is $\exp(-E_i - E_j)$ in both directions.)

1980 *Mathematics Subject Classification* (1985 Revision). Primary 93E05; Secondary 60J15.

Key words and phrases. Automatic control, associative memory, finite automata.

*This author's research was supported by the Hungarian National Foundation for Scientific Research Grant No. 1905.

Akadémiai Kiadó, Budapest

In case $N = 2^k$ the energy function E of states may be given by a quadratic form $x^T W x$, where x is a k-dimensional ± 1 vector and W is a $k \times k$ symmetric real matrix. The edges of the graph connect vertices which differ only in one coordinate. This model is called associative memory and the "association" matrix W may be used to store the system of its stable states.

(A state is stable if the energy has a local minimum there.) If an i.i.d. sequence of k-dimensional binary vectors is generated by the stationary distribution of an associative memory then we would like to reconstruct the matrix W from the sequence in an economic way. One idea for this may be to calculate the covariances. Sometimes we are satisfied by computing their signs, and this is the situation which suggested the problem discussed in this note. The estimator of a given element in the covariance matrix is a scalar product of two ± 1 vectors, here we simply use the sum of independent ± 1 bits. An automaton attempts to memorize the sign of that sum, and the natural prediction for that sign is the conditional expectation of the sign conditioned under the present state of the automaton. This way, the problem is reformulated as a control problem.

For related statistical problems, see Robbins [2], Wagner [4], and Cover-Freedman–Hellman [1].

2. Optimal control

We are given an automaton with m states. In the t-th step the input is u_t and the state is x_t. Here u_t is independent of the past $u_1, x_1, \ldots, u_{t-1}, x_{t-1}$, and

$$P(u_t = 1) = P(u_t = -1) = 1/2.$$

In case $u_t = 1$ the system is controlled by the stochastic matrix P_t:

$$P_t(i, j) = P(x_t = j \mid x_{t-1} = i, u_t = 1)$$

and in case $u_t = -1$ the control is given by Q_t:

$$Q_t(i, j) = P(x_t = j \mid x_{t-1} = i, u_t = -1).$$

Writing $S_n = u_1 + \ldots + u_n$, the conditional expectation

$$K_n = E(S_n \mid x_n)$$

depends on the sequence of control matrices

$$\mathcal{C}_n = (P_1, Q_1, P_2, Q_2, \ldots, P_{n-1}, Q_{n-1})$$

(we will use $x_0 = 1$).

We would like to maximize the expected payoff $E|K_n|$ in $\mathcal{C}_n$:

$$C(n,m) = \sup_{\mathcal{C}_n} E|K_n|.$$

Let T_n be the sign of K_n:

$$T_n = \begin{cases} +1 & \text{if } K_n \geqq 0, \\ -1 & \text{if } K_n < 0. \end{cases}$$

The payoff is $T_n S_n$, and $E|K_n| = ET_n S_n$.

EXAMPLE. Let $m = 2$, and let us use the values ± 1 (rather than 1 and 2) to code the states of x_t. We define the updating rules:

1) If $u_t = x_{t-1}$ then $x_t = x_{t-1}$.

2) If $u_t = -x_{t-1}$ then $x_t = -x_{t-1}$ with probability $p_t = 1/t$.

In other words,

$$P_t = \begin{pmatrix} 1 & 0 \\ 1/t & 1-1/t \end{pmatrix}, \quad Q_t = \begin{pmatrix} 1-1/t & 1/t \\ 0 & 1 \end{pmatrix}.$$

Then, $Eu_i x_t = 1/t$ for all i, $1 \leqq i \leqq t$, and $E|K_t| = \left| \sum_{i=1}^{t} Eu_i x_t \right| = 1$.

REMARK. Santosh Venkatesh [3] showed that among all rules of the above type (with possibly different probabilities p_t), the above choice $p_t = 1/t$ is optimal in that it maximizes the minimal covariance $\min_{1 \leqq i \leqq t} |Eu_i x_t|$.

Misinterpreting a classical lemma of Wald one may expect $C(n,m) \leqq 1$. The next lemma shows that this is not the case.

LEMMA 1. *$C(3,2) \geqq 5/4$, and $\sup_n C(n,2) \geqq 4/3$.*

We will use the following notations

$$p_t(i) = P(x_t = i), \quad e_t(i) = P(x_t = i)E(S_t \mid x_t = i) = \int_{\{x_t = i\}} S_t.$$

Thus,

$$\sum_{i=1}^{m} e_t(i) = ES_t = 0 \quad \text{and} \quad \sum_{i=1}^{m} |e_t(i)| = E|K_t|.$$

REMARK. Our strategy will be deterministic; the automaton mechanically follows the input signals. Since the payoff function is a multilinear form of the control parameters, all control parameters *must* take the extreme values 0 or 1 in the optimal solution, that is, the automaton must be deterministic.

Just as in the first example, we code the states by ± 1.

First put $x_1 = u_1$. Next, set $x_2 = 1$ for $u_2 = x_1 = 1$, and set $x_2 = -1$ otherwise. Finally, let $x_3 = -1$ for $u_3 = x_2 = -1$, and let $x_3 = 1$ otherwise. (This corresponds to the control matrices

$$P_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad Q_1 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$

$$P_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \; Q_2 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \; P_3 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \; Q_3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and the usual initial condition $x_0 = 1$, that is, $\pi_0 = (1, 0)$, $e_0 = (0, 0)$.)

Thus, for inputs $(-1, -1, -1), (-1, 1, -1), (1, -1, -1)$, the final state will be $x_3 = -1$, and it will be $x_3 = 1$ otherwise.

Consequently, $E(s_3 \mid x_3 = 1) = 1$, $E(s_3 \mid x_3 = -1) = -5/3$, and $E|K_3| = 5/4$.

To improve the bound $5/4$ to $4/3$, we repeat the above two choices for P_2, Q_2, P_3 alternately. This leads to

$$e_t \approx \frac{1}{2}[e_{t-1} + (2/3, -2/3)]$$

whence $\lim e_t = (2/3, -2/3)$ and $\lim E|K_t| = 4/3$.

It is not hard to see that the last example is optimal, and we have the exact optimum

$$C(n, 2) = \frac{4}{3}\left(1 - 4^{-\lceil n/2 \rceil}\right) \leqq \frac{4}{3}.$$

For the sake of simplicity, we only prove the weaker upper bound 2, for we only want to emphasize the drastic difference between the cases $m = 2$ and $m = 3$.

LEMMA 2. *For all* $n, C(n, 2) \leqq 2$.

The following claim clearly proves the lemma.

CLAIM. *Let* $m = 2$. *Then,* $|e_t(i)| = |\int S_t \chi(x_t = i)| \leqq 1$ *for* $i = 1, 2$.

We use induction on t. For $t = 1$ the claim is trivial. For $t > 1$, we have

$$p_t^* = p_{t-1}^* \frac{P_t + Q_t}{2}, \quad e_t^* = e_{t-1}^* \frac{P_t + Q_t}{2} + p_{t-1}^* \frac{P_t - Q_t}{2}. \tag{1}$$

(Here we used p^* to indicate row-vectors, but in the following we will be sloppier, and drop the sign * of transpose.)

Write $e_{t-1} = (a, -a)$, $p_{t-1} = (p, 1-p)$, where $|a| \leqq 1$, $0 \leqq p \leqq 1$. Then,

$$|e_t(1)| \leqq \frac{1}{2}[|a + p|^+ + |1 - a - p|^+ + |a - p|^+ + |-1 - a + p|^+] \leqq 1$$

where $|x|^+ = \max\{x, 0\}$.

LEMMA 3. *$C(n,3) \geqq \kappa \log n$, where κ is a positive constant.*

REMARK. The fact that a 3-state memory is infinitely more powerful than a 2-state memory rhymes with similar phenomena observed in the statistical literature.

PROOF. Starting with $p_0(1) = 1$, $p_0(i) = 0$ for $i > 1$, we have $e_0(i) = 0$ for $i \geqq 1$. Using the recursion (1), our task is to maximize

$$E|K_t| = \sum_{i=1}^{m} |e_t(i)|,$$

or at least to show that it tends to infinity with rate $\log t$ if $m \geqq 3$. For this aim we will define the control strategies P_t, Q_t in such a way that if $E|K_t| = a$ then $E|K_{t+v}| \geqq a + 1$ holds true with $v \leqq a2^a$.

In our construction the states 1 and 3 will be "large collectors" with opposite signs, while the state 2 will have an auxiliary character. At the beginning of the recursion step, we gather everything in large collectors: $e_t(1) = -a/2$, $e_t(3) = a/2$. We may suppose that $p_t(3) \geqq p_t(1)$. Thus $E(S_t \mid x_t = 3) \leqq a$. First we push everything from state 3 to state 2. The conditional expectation here will be drained step by step walking from 2 to 3 with $u_t = 1$ and remaining in 2 with $u_t = -1$.

In one step, the conditional expectation decreases with 1, thus in less than a steps the modulus of the expected value at state 2 will be less than 1. With an appropriate last step we can change it to 0. In the meantime the probabilities will decrease to 2^{-a}, yielding 2^{-a} gain in a steps. Repeating the whole procedure 2^a times, the desired increase is achieved.

3. Open problems

As the reader can see, we only have preliminary results. We do not know the optimal strategy or even the order of magnitude of $C(n,m)$ for $m \geqq 3$.

Standard subadditivity arguments give a lower bound

$$C(n,m) > (\kappa \log n/M)^{\lfloor M \rfloor}$$

where $M = \log m/\log 3$. We believe this to be close to the true order of magnitude, but cannot prove it. In fact, we do not have any reasonable upper bounds on $C(n,m)$, $m > 2$.

REFERENCES

[1] COVER, T. M., FREEDMAN, M. A. and HELLMAN, M. E., Optimal finite memory learning algorithms for the finite sample problem, *Information and Control* **30** (1976), 49–85. *MR* **53** #9719

[2] ROBBINS, H., Sequential decision problem with a finite memory, *Proc. Nat. Acad. Sci. U.S.A.* **42** (1956), 920–923. *MR* **18**-606

[3] VENKATESH, SANTOSH, 1989 (personal communication).
[4] WAGNER, T. J., Estimation of the mean with time-varying finite memory, *IEEE Trans. Information Theory* **IT-18** (1972), 523–525.

(Received August 28, 1989)

DEPARTMENT OF MATHEMATICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NJ 08903
U.S.A.

DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF DELAWARE
NEWARK, DE 19716
U.S.A.

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 179–181

RESTRICTIONS OF ADJOINT OPERATORS IN HILBERT SPACE

Z. SEBESTYÉN

The characterization problem for restrictions of positive bounded operators on Hilbert space to (subsets or to) linear subspaces of the ground Hilbert space is given in [3, Theorem] (see further [4], [5]). The same problem for positive self-adjoint and not necessarily bounded operators is solved in [4, Theorem 1].

The aim of this note is first to characterize restrictions of adjoint operators of densely defined linear operators to linear and not necessarily closed subspace of the ground Hilbert space. Our necessary and sufficient condition (ii) in Theorem below is a counterpart of (ii) in [4, Theorem 1] for the positive self-adjoint case.

The characterization problem of restrictions of self-adjoint operators to not necessarily dense linear subspace of the ground Hilbert space remains still open.

Our approach is elementary in the sense that it goes back to the definition of an adjoint operator. One resembles also the range characterization of adjoint operators in Hilbert space given in [2, Theorem 1].

As a second corollary we give a factorization result generalizing earlier ones in [1] and [2].

Let E be a given linear operator defined on a linear subspace D of a (complex) Hilbert space H with values in H. It is natural to ask the following question: under what condition does there exist a densely defined, not necessarily bounded linear operator A in the Hilbert space H such that the adjoint operator A^* extends E. This means in other words that $D(A)$, the domain of A, contains D and at the same time that

$$A^*x = Ex \quad \text{holds for each } x \text{ from } D. \tag{1}$$

The answer is contained in the following

THEOREM. *Let E be a linear operator defined on a linear subspace D of a Hilbert space H. The following two statements are equivalent:*

(i) *There exists a densely defined operator A in H satisfying* (1);

(ii) $K := [y \in H : \sup\{|(Ex, y)| : x \in D, \|x\| \leqq 1\} < \infty]$ *is dense in H.*

PROOF. Assume first (i). Then the domain $D(A)$ of A is dense in H. Hence the inclusion $D(A) \subset K$ proves (ii); indeed, to prove that a vector y

1980 *Mathematics Subject Classification.* Primary 47A20; Secondary 47B15.
Key words and phrases. Hilbert space operators, closed operators, restrictions.

Akadémiai Kiadó, Budapest

from $D(A)$ belongs to K it is enough to use (1) for each x from D and the adjoint identity

$$(Ex, y) = (A^*x, y) = (x, Ay)$$

to get the desired majorization

$$|(Ex, y)| \leqq \|Ay\|\|x\| \quad \text{for each } x \text{ from } D \text{ and } y \text{ from } D(A).$$

Assume now that (ii) holds true. Then we define an operator A on K with values in $\bar{D}$, the closure in norm of D, as follows. Given a vector y from K we know by (ii) in Theorem that the linear functional on D with values (Ex, y) in x (from D) is bounded. Thus it has a unique continuous extension to $\bar{D}$. The celebrated Riesz representation theorem gives a unique vector Ay in (the Hilbert space) $\bar{D}$ such that

$$(Ex, y) = (x, Ay) \quad \text{holds for each } x \text{ from } D \text{ and } y \text{ from } K.$$

That the map $A: y \mapsto Ay$ so defined on K with values in H is linear, is plain. The identity (2) shows that D belongs to the domain of the adjoint A^* of A and the identity (1) as well. Hence A^* extends E or E is a restriction of A^* to D. The proof is complete.

COROLLARY 1. *For the linear operator $E: D \to H$ the following statements are equivalent:*

(i') *E is a restriction of a bounded linear operator on H;*

(ii') *$K = H$;*

(iii') *E is bounded.*

PROOF. Assume first (i'): let B be a bounded linear operator on H with restriction E to D. Then $A^* = B$ does the same if $A = B^*$ and since $H = D(A^*) \subset K$ we get (ii'). Now assuming (ii') we see that the linear functionals used in proof of the Theorem are pointwise bounded in y, $y \in K$, $\|y\| \leqq 1$ hence uniformly bounded by the Banach–Steinhaus theorem. $|(Ex, y)| \leqq m\|x\|\|y\|$ holds for each x from D and y from H, where $m \geqq 0$ is a constant independent of x and y. This shows that E is bounded, (iii') follows. It is easy to see that (iii') implies (i').

COROLLARY 2. *Let B and C be densely defined operators in H such that domain $D(B)$ of B contains domain $D(C)$ of C.*

(a) *There exists a densely defined operator A such that*

$$Cx = A^*Bx \quad \text{holds for each } x \text{ from } D(C), \tag{3}$$

(b) $K := [y \in H : \sup\{|(Cx, y)| : x \in D(C), \|Bx\| \leqq 1\} < \infty]$ *is dense in H.*

PROOF. Assuming (a) we see $D(A) \subset K$ thus (b):

$$|(Cx, y)| = |(A^*Bx, y)| = |(Bx, Ay)| \leqq \|Ay\|\|Bx\|$$

follows by (3) for each x from $D(C)$ and y from $D(A)$.

Assume now (b). For any y from K the linear functional on range of B $R(B)$, $Bx \mapsto (Cx, y)$, where x runs over $D(B)$, is continuous. Replacing D of Theorem with $R(B)$ we have an operator A on K with values in $\overline{R(B)}$ such that

$$(Cx, y) = (Bx, Ay) \text{ holds for each } x \text{ from } D(B) \text{ and } y \text{ from } K.$$

This shows that $R(B)$ belongs to domain $D(A^*)$ of A^*, the adjoint operator of A and the identity (3) as well.

COROLLARY 3. *A is bounded in Corollary 2 if and only if $K = H$ if and only if*

(c) $\|Cx\| \leqq m\|Bx\|$ *holds for each x from $D(C)$, where $m \geqq 0$ is some constant.*

REMARK. Corollaries 2 and 3 are versions of [1, Theorem 1] expressed in adjoints but for unbounded factorization, too.

REFERENCES

[1] DOUGLAS, R. G., On majorization, factorization, and range inclusion of operators on Hilbert space, *Proc. Amer. Math. Soc.* **17** (1966), 413–415. *MR* **34** #3315

[2] SEBESTYÉN, Z., On ranges of adjoint operators in Hilbert space, *Acta Sci. Math. (Szeged)* **46** (1983), 295–298. *MR* **85i**: 47003a

[3] SEBESTYÉN, Z., Restrictions of positive operators, *Acta Sci. Math. (Szeged)* **46** (1983), 299–301. *MR* **85i**: 47003b

[4] SEBESTYÉN, Z. and STOCHEL, J., Restrictions of positive self-adjoint operators, *Acta Sci. Math. (Szeged)* **55** (1991), 149–154. *MR* **92j**: 47042

[5] SEBESTYÉN, Z. and KAPOS, L., Extremal positive and self-adjoint extensions of suboperators, *Period. Math. Hungar.* **20** (1989), 75–80. *MR* **90i**: 47008

(Received August 31, 1989)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
ALKALMAZOTT ANALÍZIS TANSZÉK
MÚZEUM KRT. 6–8
H–1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 183–188

ON APPROXIMATION BY TRIGONOMETRIC POLYNOMIALS IN L_u^p-SPACES

N. X. KY

Introduction

In the present paper we investigate the best approximation by trigonometric polynomials in weighted spaces with weights satisfying the so-called A_p-condition introduced by R. Hunt, B. Muckenhoupt, R. Wheeden [5]. Such weights play an important role in solutions of many different problems of harmonic analysis, theory of operators, approximation theory (see e.g. [4], [5], [6]). In Part 1 Jackson and Bernstein type inequalities are given. As a consequence of that, we obtain an equivalent theorem, which states the relation between the order of the best approximation, the simultaneous approximation and the norm of derivatives of best approximating polynomials. In Part 2 we give the relation between the best approximation and Peetre K-functional.

1. Jackson and Bernstein inequalities

Throughout this paper, let $1 < p < \infty$ and let $c(p, \dots)$ denote a constant depending only on its variables (it may be different in different formulas).

A 2π-periodic measurable function $u(x)$ is called satisfied the A_p-condition if $0 < u(x) < \infty$ a.e. and for every finite interval I, we have

$$\left(\frac{1}{|I|}\int_I u(x)dx\right)\left(\frac{1}{|I|}\int_I u^{-1/(p-1)}(x)dx\right)^{p-1} \leqq c(p) \tag{1}$$

here $|I|$ denotes the length of I.

Let A_p be the set of all weights satisfying the A_p-condition. Such weights were introduced by R. Hunt, B. Muckenhoupt, R. Wheeden [5] for investigation of the trigonometric conjugate operator. We remark that in the case $p = 2$, those weights were considered early by H. Helson and G. Szegő [4].

1980 *Mathematics Subject Classification* (1985 Revision). Primary 42A10.
Key words and phrases. Jackson and Bernstein inequalities, A_p-condition.

Akadémiai Kiadó, Budapest

Let L_u^p be the Banach space of 2π-periodic measurable functions with the norm

$$\|f\|_{p,u} = \left\{ \int_0^{2\pi} |f(x)|^p u(x)dx \right\}^{\frac{1}{p}} \quad (< \infty).$$

In the case $u \equiv 1$, we write L^p, $\|f\|_p$ instead of L_u^p, $\|f\|_{p,u}$, respectively.

From the definition of the A_p-condition it follows that if $u \in A_p$ then it is integrable on $[0, 2\pi]$. Consequently, the space L_u^p contains every trigonometric polynomial. We can define

$$E_n(f)_{p,u} = \inf_{t_n \in T_n} \|f - t_n\|_{p,u} \qquad (f \in L_u^p,\ n = 0, 1, \dots) \tag{2}$$

where T_n denotes the set of all trigonometric polynomials of degree at most n.

We introduce the following classes of functions: $M_{p,u}^{(o)} = L_u^p$, and for $k = 1, 2, \dots$, $M_{p,u}^{(k)}$ consists of all functions f having the property that $f, f', \dots, f^{(k-1)}$ are absolutely continuous on $[0, 2\pi]$, $f^{(k)} \in L_u^p$.

The following theorem is true.

THEOREM 1. *Let $u \in A_p$ and let $k \geq 1$ be an integer. For every $f \in M_{p,u}^{(k)}$ and $n = 1, 2, \dots$ we have*

$$E_n(f)_{p,u} \leq \frac{c(k,p,u)}{n^k} \|f^{(k)}\|_{p,u}. \tag{3}$$

For the proof of the theorem we need

LEMMA 1. *Let $u \in A_p$. If $f \in L_u^p$ then it is integrable on $[0, 2\pi]$ and*

$$\|f\|_1 \leq c(p,u)\|f\|_{p,u}. \tag{4}$$

PROOF. Let $I := [0, 2\pi]$. Since $u \in A_p$ we have

$$\int_I u(x)dx > 0$$

therefore from (1) it follows that

$$\int_I u^{-1/(p-1)}(x)dx < \infty.$$

Let now $\frac{1}{p} + \frac{1}{q} = 1$. Then f can be written in the form:

$$f = (fu^{\frac{1}{p}})(u^{-1/(p-1)})^{\frac{1}{q}} =: gh.$$

Since $g \in L^p$, $h \in L^q$ we have $f \in L^1$ and (3) follows by Hölder's inequality.

Let now $\varphi \in L^p_u$. Since $\varphi \in L^1$ it has the trigonometric conjugate function which will be denoted usually by $\tilde{\varphi}$. In [5] the authors proved that

$$\|\tilde{\varphi}\|_{p,u} \leqq c(p,u)\|\varphi\|_{p,u}. \tag{5}$$

On the other hand, the function φ has the Fourier-series

$$\varphi(x) \sim \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k \cos kx + b_k \sin kx). \tag{6}$$

Denote by $\sigma_n(\varphi)$ the n-th Fejér mean of series (6). From Theorem 8 of [5] we get

$$\|\sigma_n(\varphi)\|_{p,u} \leqq c(p,u)\|\varphi\|_{p,u} \qquad (n = 0, 1, \dots). \tag{7}$$

Consequently, by the Banach–Steinhaus theorem

$$\|\sigma_n(\varphi) - \varphi\|_{p,u} \to 0 \qquad (n \to \infty).$$

So for every $f \in L^p_u$

$$E_n(f)_{p,u} \to 0 \qquad (n \to \infty).$$

Proof of Theorem 1. It is enough to see (3) for $k = 1$. The cases $k \geqq 2$ then can be obtained by induction. Let $f \in M^{(1)}_{p,u}$ and

$$f(x) \sim \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k \cos kx + b_k \sin kx) \tag{8}$$

then

$$\tilde{f}(x) \sim \sum_{k=1}^{\infty}(b_k \cos kx - a_k \sin kx). \tag{9}$$

Since f is absolutely continuous, it is easy to see that

$$f'(x) \sim \sum_{k=1}^{\infty} k(b_k \cos kx - a_k \sin kx). \tag{10}$$

Furthermore by (7) we have

$$\|\sigma_n(f')\|_{p,u} \leqq c(p,u)\|f'\|_{p,u} \qquad (n = 0, 1, \dots). \tag{11}$$

Now, by (11) using Lemma 1 of [1] for two series (9) and (10) we get

$$\|\sigma_n(\tilde{f}) - (\tilde{f})\|_{p,u} \leqq \frac{c(p,u)}{n}\|f'\|_{p,u} \qquad (n = 1, 2, \dots). \tag{12}$$

Using twice (5) for the function $\varphi = \sigma_n(f) - f$ we obtain

$$\|\sigma_n(f) - f\|_{p,u} \leqq c(p,u)\|\sigma_n(\bar{f}) - \tilde{f}\|_{p,u} \qquad (n = 1,2,\dots)$$

which together with (12) proves that

$$\|\sigma_n(f) - f\|_{p,u} \leqq \frac{c(p,u)}{n}\|f'\|_{p,u} \qquad (n = 1,2,\dots)$$

and the last inequality implies (3) for $k = 1$.

THEOREM 2. *We have for every* $t_n \in T_n$ $(n = 1,2,\dots)$

$$\|t_n'\|_{p,u} \leqq c(p,u)n\|t_n\|_{p,u}. \tag{13}$$

PROOF. It is enough to see (13) for any $t_n \in T_n$ for which $\|t_n\|_{p,u} = 1$. Let

$$t_n(x) = \frac{a_0}{2} + \sum_{k=1}^{n}(a_k \cos kx + b_k \sin kx).$$

Then

$$t_n'(x) = \sum_{k=1}^{n} k(b_k \cos kx - a_k \sin kx),$$

$$\tilde{t}_n(x) = \sum_{k=1}^{n}(b_k \cos kx - a_k \sin kx).$$

Using (5) and (7) we have

$$\|\sigma_n(\tilde{t}_n/n)\|_{p,u} \leqq c(p,u)\frac{1}{n}\|\tilde{t}_n\|_{p,u} \leqq \frac{c(p,u)}{n}\|t_n\|_{p,u} \leqq \frac{c(p,u)}{n}.$$

Therefore by application of Lemma 2 in [2] for the polynomials t_n' and $\tilde{t}_n$ as two series we get

$$\|\sigma_n(t_n'/n)\|_{p,u} \leqq c(p,u). \tag{14}$$

By the same way we have also

$$\|\sigma_{2n}(t_n'/n)\|_{p,u} \leqq c(p,u). \tag{15}$$

Let now $\theta_n = 2\sigma_{2n} - \sigma_n$ be the de la Vallée–Poussin operator. Then from (14) and (15) it follows that

$$\|\theta_n(t_n'/n)\|_{p,u} \leqq c(p,u). \tag{16}$$

Finally, since $\theta_n(t_n'/n) = t_n'/n$ we have by (16)

$$\|t_n'\|_{p,u} \leqq c(p,u)n = c(p,u)n\|t_n\|_{p,u},$$

which was to be proved.

2. An equivalent theorem

Peetre K-functional between L_u^p and $M_{p,u}^{(k)}$ is defined by the formula

$$K_k(L_u^p, f, t) = \inf_{g \in M_{p,u}^{(k)}} \{\|f-g\|_{p,u} + t\|g^{(k)}\|_{p,u}\} \tag{17}$$
$$(f \in L_u^p,\ 0 < t < \infty,\ k = 1, 2, \dots).$$

Let furthermore $t_n(f)$ be the n-th polynomial of best approximation of f in L_u^p. By virtue of Theorem 3 in [3], using Theorems 1 and 2 we get

THEOREM 3. *Let j, k, r be positive integers, $\alpha > 0$, and $1 \leqq s \leqq \infty$. Let $f \in L_u^p$. The following statements are equivalent:*

(a) $\left\{\sum_{n=1}^{\infty} [n^{r+\alpha} E_n(f)]^s \frac{1}{n}\right\}^{\frac{1}{s}} < \infty$

(b) $f \in M_{p,u}^{(k)}$ *and* $\left\{\sum_{n=1}^{\infty} \left[n^{r+\alpha-k} \|t_n^{(k)}(f) - f^{(k)}\|_{p,u}\right]^s \frac{1}{n}\right\}^{\frac{1}{s}} < \infty \quad (k < r+\alpha)$

(c) $\left\{\sum_{n=1}^{\infty} \left[n^{r+\alpha-j} \|t_n^{(j)}(f)\|_{p,u}\right]^s \frac{1}{n}\right\}^{\frac{1}{s}} < \infty \quad (r+\alpha < j)$

(d) $\left\{\int_0^{\infty} \left[t^{-(r+\alpha)/j} K_j(L_u^p, f, t)\right]^s \frac{dt}{t}\right\} < \infty.$

3. A direct and converse theorem

The relation between $E_n(f)_{p,u}$ and $K_k(L_u^p, f, t)$ can be written in the following stronger form.

THEOREM 4. *Let $k \geqq 1$ be an integer. For any $f \in L_u^p$ we have*

$$E_n(f)_{p,u} \leqq c(p,u) K_k\left(L_u^p, f, \frac{1}{n}\right) \quad (n \geqq k) \tag{18}$$

$$K_k\left(L_u^p, f, \frac{1}{n}\right) \leqq \frac{c(p,u)}{n^k} \sum_{k=1}^{n} j^{k-1} E_j(f)_{p,u} \quad (n \geqq 1). \tag{19}$$

The idea of the proof of Theorem 4 is known. It will be based on two inequalities (3) and (13). The analogue of the proof can be found for example in [6], therefore here is omitted.

REMARK. One of the most important problems in approximation theory is to give estimates of the best approximation in terms of some constructive properties of functions, for example in terms of moduli of continuity. In the case of the best approximation (2), this problem is still open.

REFERENCES

[1] ALEXITS, G., Sur l'ordre de grandeur de l'approximation d'une fonction par les moyennes de sa série de Fourier, *Mat. Fiz. Lapok* **48** (1941), 410–422. *MR* **8**–261

[2] ALEXITS, G., Sur l'ordre de grandeur de l'approximation d'une fonction périodique par les sommes de Fejér, *Acta Math. Acad. Sci. Hungar.* **3** (1952), 29–42. *MR* **14**–370

[3] BUTZER, P. L. and SCHERER, K., On the fundamental approximation theorems of D. Jackson, S. N. Bernstein and theorems of M. Zamansky and S. B. Stečkin, *Aequationes Math.* **3** (1969), 170–185. *MR* **41** #8897

[4] HELSON, H. and SZEGŐ, G., A problem in prediction theory, *Ann. Mat. Pura Appl.* (4) **51** (1960), 107–138. *MR* **22** #12343

[5] HUNT, R., MUCKENHOUPT, B. and WHEEDEN, R. L., Weighted norm inequalities for the conjugate function and Hilbert transform, *Trans. Amer. Math. Soc.* **176** (1973), 227–251. *MR* **47** #701

[6] JOÓ, I., On Riesz bases, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **31** (1988), 141–153. *MR* **90i**: 42045

[7] KY, N. X., On Jackson and Bernstein type approximation theorems in the case of approximation by algebraic polynomials in L_p-space, *Studia Sci. Math. Hungar.* **9** (1974), 405–415. *MR* **53** #6182

(Received September 4, 1989)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
EGYETEMI SZÁMÍTÓKÖZPONT
BOGDÁNFY U. 10/B
H–1117 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 189–196

ON THE FOURIER TRANSFORM OF THE MODIFIED BESSEL FUNCTION WITH RESPECT TO THE ORDER

T. FÉNYES

In [1] Cooke proved that for $|\arg z| < \frac{\pi}{2}$

$$(1)\quad \int_0^\infty I_\xi(z)\cos\Phi\xi d\xi = \frac{e^{z\cos\Phi}}{2} - \frac{1}{2\pi}\int_0^\infty \left\{ \frac{\pi+\Phi}{(\pi+\Phi)^2+t^2} + \frac{\pi-\Phi}{(\pi-\Phi)^2+t^2} \right\} e^{-z\,\mathrm{ch}\,t}dt$$

if $|\Phi| < \pi$. The first term is replaced by $\frac{e^{-z}}{4}$ if $|\Phi| = \pi$, and by zero otherwise. In particular, for $\Phi = 0$,

$$(2)\quad \int_0^\infty I_\xi(z)d\xi = \frac{e^z}{2} - \int_0^\infty \frac{e^{-z\,\mathrm{ch}\,t}}{\pi^2+t^2}dt.$$

In this paper we calculate the integrals

$$(3)\quad \int_0^\infty I_\xi(t)\sin\Phi\xi d\xi, \quad t \geqq 0,$$

$$(4)\quad \int_0^\infty \xi I_\xi(t)d\xi, \quad t \geqq 0.$$

We shall apply the above integrals to the solution of an integral equation of convolutional type in connection with a heat-conduction problem.

We start with the known formula

$$(5)\quad I_\xi(t) = \frac{1}{\pi}\int_0^\pi e^{t\cos\Theta}\cos\Theta\xi d\Theta - \frac{\sin\xi\pi}{\pi}\int_0^\infty e^{-t\,\mathrm{ch}\,u-\xi u}du$$

(see Watson [2], p. 181).

1980 *Mathematics Subject Classification.* Primary 33C10; Secondary 35K05, 44A10.
Key words and phrases. Bessel functions, operational calculus, fractional calculus.

[1]Research (partially) supported by the National Foundation for Scientific Research Grant No. 6032/6319.

So we have

$$\int_0^\infty I_\xi(t) \sin \Phi \xi d\xi = g_1(\Phi) + g_2(\Phi), \tag{6}$$

where

$$\begin{aligned} g_1(\Phi) &= \frac{1}{\pi} \int_0^\infty \sin \xi\Phi \int_0^\pi e^{t \cos \Theta} \cos \Theta \xi d\Theta d\xi \\ g_2(\Phi) &= -\frac{1}{\pi} \int_0^\infty \sin \xi\Phi \sin \xi\pi \int_0^\infty e^{-t \operatorname{ch} u - \xi u} du\, d\xi \end{aligned} \qquad |\Phi| \neq \pi. \tag{7}$$

Though $g_1(\Phi)$, $g_2(\Phi)$ have singularities if $\Phi = \pm\pi$, we show that their sum is continuous for $\Phi = \pm\pi$. By the legitimate inversion of the order of integration in $g_2(\Phi)$ we obtain

$$\begin{aligned} g_2(\Phi) &= -\frac{1}{2\pi} \int_0^\infty \int_0^\infty \cos(\Phi - \pi) \xi e^{-\xi u - t \operatorname{ch} u} d\xi\, du + \\ &+ \frac{1}{2\pi} \int_0^\infty \int_0^\infty \cos(\Phi + \pi) \xi e^{-\xi u - t \operatorname{ch} u} d\xi\, du = \\ &= -\frac{1}{2\pi} \int_0^\infty \left(\frac{u}{u^2 + (\Phi - \pi)^2} - \frac{u}{u^2 + (\Phi + \pi)^2} \right) e^{-t \operatorname{ch} u} du. \end{aligned} \tag{8}$$

Integration by parts gives

$$g_2(\Phi) = \frac{1}{2\pi} \log \left| \frac{\pi - \Phi}{\pi + \Phi} \right| e^{-t} - \frac{t}{4\pi} \int_0^\infty \log \frac{u^2 + (\Phi - \pi)^2}{u^2 + (\Phi + \pi)^2} \operatorname{sh} u e^{-t \operatorname{ch} u} du. \tag{9}$$

Here the integral does not exist for $t = 0$, however, the whole second term tends to zero if $t \to 0$.

Applying the Fourier series expansion

$$e^{t \cos \Theta} = I_0(t) + 2 \sum_{n=1}^\infty I_n(t) \cos n\Theta$$

(see Luke [3]) we can write

$$g_1(\Phi)=\frac{1}{\pi}\int_0^{\infty}\sin\xi\Phi\int_0^{\pi}\Big[I_0(t)+2\sum_{n=1}^{\infty}I_n(t)\cos n\Theta\Big]\cos\Theta\xi\,d\Theta\,d\xi=$$

(10)

$$=\frac{I_0(t)}{\pi}\int_0^{\infty}\frac{\sin\xi\Phi\sin\xi\pi}{\xi}d\xi+\frac{2}{\pi}\sum_{n=1}^{\infty}I_n(t)\int_0^{\infty}\sin\xi\Phi\int_0^{\pi}\cos n\Theta\cos\Theta\xi\,d\Theta\,d\xi=$$

$$=\frac{I_0(t)}{2\pi}\log\left|\frac{\pi+\Phi}{\pi-\Phi}\right|+\frac{1}{\pi}\sum_{n=1}^{\infty}I_n(t)\int_0^{\infty}\Big(\frac{\sin(\xi+n)\pi}{\xi+n}\sin\xi\Phi+\frac{\sin(\xi-n)\pi}{\xi-n}\sin\xi\Phi\Big)d\xi.$$

Here the inversion of the order of integrations and summation is legitimate, which can easily be seen. We must evaluate

$$I=\int_0^{\infty}\Big(\frac{\sin(\xi+n)\pi}{\xi+n}\sin\xi\Phi+\frac{\sin(\xi-n)\pi}{\xi-n}\sin\xi\Phi\Big)d\xi. \tag{11}$$

Applying elementary trigonometrical additional formulas, and introducing the substitutions $\xi+n=u$, $\xi-n=u$, respectively, after some routine steps (10) can be reduced to the form

$$I=\frac{1}{2}\cos\Phi n\Big(\int_{-n}^{\infty}\frac{\cos(\pi-\Phi)u-\cos(\pi+\Phi)u}{u}du+\int_{n}^{\infty}\frac{\cos(\pi-\Phi)u-\cos(\pi+\Phi)u}{u}du\Big)+$$

$$+\frac{1}{2}\sin\Phi n\Big(\int_{-n}^{\infty}\frac{\sin(\pi-\Phi)u+\sin(\pi+\Phi)u}{u}du-\int_{n}^{\infty}\frac{\sin(\pi-\Phi)u+\sin(\pi+\Phi)u}{u}du\Big)=$$

(12)

$$=\cos\Phi n\Big[\int_0^{\infty}\frac{\cos(\pi-\Phi)u-\cos(\pi+\Phi)u}{u}du+\int_0^{n}\frac{-\cos(\pi-\Phi)u+\cos(\pi+\Phi)u}{u}du\Big]+$$

$$+\sin\Phi n\int_0^{n}\frac{\sin(\pi-\Phi)u+\sin(\pi+\Phi)u}{u}du=$$

$$=\sin\Phi n\,[S_1[(\pi+\Phi)n]+S_1[(\pi-\Phi)n]]+\cos\Phi n\log\left|\frac{\pi+\Phi}{\pi-\Phi}\right|+$$

$$+\cos\Phi n\,[S_1[(\pi-\Phi)n]-S_1[(\pi+\Phi)n]],$$

where

$$S_i(x)=\int_0^x \frac{\sin u}{u}du, \qquad S_1(x)=\int_0^x \frac{1-\cos u}{u}du.$$

So by (10), (12) we have

$$g_1(\Phi)=\frac{I_0(t)}{2\pi}\log\left|\frac{\pi+\Phi}{\pi-\Phi}\right|+\frac{1}{\pi}\log\left|\frac{\pi+\Phi}{\pi-\Phi}\right|\sum_{n=1}^{\infty}I_n(t)\cos\Phi n+$$

$$+\frac{1}{\pi}\sum_{n=1}^{\infty}I_n(t)\cos\Phi n\,[S_1[(\pi-\Phi)n]-S_1[(\pi+\Phi)n]]+$$

$$+\frac{1}{\pi}\sum_{n=1}^{\infty}I_n(t)\sin\Phi n[S_i[(\pi+\Phi)n]+S_i[(\pi-\Phi)n]]= \tag{13}$$

$$=\frac{1}{2\pi}\log\left|\frac{\pi+\Phi}{\pi-\Phi}\right|e^{t\cos\Phi}+\frac{1}{\pi}\sum_{n=1}^{\infty}I_n(t)\cos\Phi n(S_1[(\pi-\Phi)n]-S_1[(\pi+\Phi)n])+$$

$$+\frac{1}{\pi}\sum_{n=1}^{\infty}I_n(t)\sin\Phi n(S_i[(\pi+\Phi)n]+S_i[(\pi-\Phi)n]).$$

It is easily seen that for fixed t the infinite series occurring in (13) converge uniformly with respect to Φ.

Summing (9) and (13) we have

$$\int_0^{\infty}I_\xi(t)\sin\Phi\xi\,d\xi=\frac{1}{2\pi}\log\left|\frac{\pi+\Phi}{\pi-\Phi}\right|(e^{t\cos\Phi}-e^{-t})-$$

$$-\frac{t}{4\pi}\int_0^{\infty}\log\frac{u^2+(\Phi-\pi)^2}{u^2+(\Phi+\pi)^2}\,\mathrm{sh}\,u\,e^{-t\,\mathrm{ch}\,u}du+ \tag{14}$$

$$+\frac{1}{\pi}\sum_{n=1}^{\infty}I_n(t)\cos\Phi n(S_1[(\pi-\Phi)n]-S_1[(\pi+\Phi)n])+$$

$$+\frac{1}{\pi}\sum_{n=1}^{\infty}I_n(t)\sin\Phi n(S_i[(\pi+\Phi)n]+S_i[(\pi-\Phi)n]).$$

Since

$$\log\left|\frac{\pi+\Phi}{\pi-\Phi}\right|(e^{t\cos\Phi}-e^{-t})$$

tends to zero for $|\Phi| \to \pi$, applying Fourier's theorem we get that (14) also holds for $|\Phi| = \pi$. On the other hand, we can see that (14) is more complicated than (1).

Let us now consider the integral

$$\int_0^\infty \xi \cos \Phi \xi I_\xi(t) d\xi, \tag{15}$$

which obviously converges uniformly, since

$$\left| \int_0^\infty \xi \cos \Phi \xi I_\xi(t) d\xi \right| \leqq \int_0^\infty \xi I_\xi(t) d\xi < \int_0^1 I_\xi(t) d\xi + \int_1^\infty \xi I_\xi(t) d\xi,$$

and applying the recurrence relation

$$I_{\xi-1}(t) - I_{\xi+1}(t) = \frac{2\xi}{t} I_\xi(t) \tag{16}$$

we have

$$\int_0^\infty \xi I_\xi(t) d\xi < \int_0^1 I_\xi(t) d\xi + \frac{t}{2} \int_1^\infty (I_{\xi-1}(t) - I_{\xi+1}(t)) d\xi =$$

$$= \int_0^1 I_\xi(t) d\xi + \frac{t}{2} \left(\int_0^\infty I_\xi(t) d\xi - \int_2^\infty I_\xi(t) d\xi \right) = \int_0^1 I_\xi(t) d\xi + \frac{t}{2} \int_0^2 I_\xi(t) d\xi.$$

So by an elementary property of the Fourier transformation with (14) we obtain

$$\int_0^\infty \xi I_\xi(t) d\xi = \left[\frac{d}{d\Phi} \int_0^\infty I_\xi(t) \sin \Phi \xi d\xi \right]_{\Phi=0} =$$

$$= \frac{1}{\pi^2}(e^t - e^{-t}) + t \int_0^\infty \frac{1}{u^2 + \pi^2} \operatorname{sh} u \, e^{-t \operatorname{ch} u} du + \tag{17}$$

$$+ \frac{2}{\pi} \sum_{n=1}^\infty n I_n(t) S_i(\pi n) + \frac{2}{\pi^2} \sum_{n=1}^\infty I_n(t)(\cos \pi n - 1).$$

Here the differentiation under the integral sign is legitimate for every fixed $t > 0$ and so is the termwise differentiation of the infinite series in (14) since the obtained integral and infinite series are uniformly convergent, as it can easily

be seen by a simple majorization and application of the above recurrence relation (16). Moreover,

$$t\int_0^\infty \frac{1}{u^2+\pi^2}\,\mathrm{sh}\,u\,e^{-t\,\mathrm{ch}\,u}du$$

tends to zero for $t \to 0$, since by integration by parts we have

$$t\int_0^\infty \frac{\mathrm{sh}\,u}{u^2+\pi^2}e^{-t\,\mathrm{ch}\,u}du = \frac{e^{-t}}{\pi^2} - 2\int_0^\infty \frac{u}{(\pi^2+u^2)^2}e^{-t\,\mathrm{ch}\,u}du,$$

taking the value zero for $t=0$. So it is

$$\int_0^\infty \xi I_\xi(t)d\xi = \frac{e^t}{\pi^2} - 2\int_0^\infty \frac{u}{(\pi^2+u^2)^2}e^{-t\,\mathrm{ch}\,u}du + \frac{2}{\pi^2}\sum_{n=1}^\infty (-1)^n I_n(t) -$$

$$-\frac{2}{\pi^2}\sum_{n=1}^\infty I_n(t) + \frac{2}{\pi}\sum_{n=1}^\infty n I_n(t) S_i(\pi n).$$

Substituting

$$e^{-t} = I_0(t) + 2\sum_{n=1}^\infty (-1)^n I_n(t)$$

$$e^t = I_0(t) + 2\sum_{n=1}^\infty I_n(t)$$

in the above formula we obtain the final result

$$\int_0^\infty \xi I_\xi(t)d\xi = \frac{e^{-t}}{\pi^2} - 2\int_0^\infty \frac{u}{(u^2+\pi^2)^2}e^{-t\,\mathrm{ch}\,u}du + \frac{2}{\pi}\sum_{n=1}^\infty n S_i(\pi n) I_n(t). \tag{18}$$

(18) is very convenient for numerical calculations. It is easy to obtain a numerical value for the integral for any fixed t. On the other hand, the functions $S_i(\pi n)$, $I_n(t)$ are tabulated (see Abramowitz–Stegun [4], Luke [3]).

An application of (18) *to a heat-conduction problem*

Let us solve the following integral equation:

$$\int_0^t f(\tau)K_0(t-\tau)d\tau = 1, \qquad t>0, \tag{19}$$

which occurs if we use the fractional calculus to the solution of the cylindrical heat equation

$$\frac{\partial \vartheta(r,t)}{\partial r^2} + \frac{1}{r}\frac{\partial \vartheta(r,t)}{\partial r} = \frac{1}{a}\frac{\partial \vartheta(r,t)}{\partial t}.$$

If we denote the Laplace transform of $f(t)$ by $F(p)$, we have

$$\mathcal{L}[f(t)] = F(p) = \frac{\sqrt{p^2-1}}{p\log\left(p+\sqrt{p^2-1}\right)} \quad \text{(see for example [4])}. \tag{20}$$

Moreover, (20) can be written as

$$F(p) = \frac{\sqrt{p^2-1}-p}{p\log\left(p+\sqrt{p^2-1}\right)} + \frac{1}{\log\left(p+\sqrt{p^2-1}\right)}.$$

Here the inverse transform of $\dfrac{\sqrt{p^2-1}-p}{p}$ is

$$\mathcal{L}^{-1}\left[\frac{\sqrt{p^2-1}-p}{p}\right] = -\int_0^t \frac{I_1(u)}{u}du.$$

For the determination of the inverse transform of the logarithmic term, we make use of the idea of the papers Rühs [5], Schaar [6], in which the authors applied Mikusinski's operational calculus in solving singular integral equations.

Obviously

$$\frac{1}{p} = \int_0^\infty e^{-pu}du, \quad \mathrm{Re}[p] > 0,$$

so we have

$$\frac{1}{\log\left(p+\sqrt{p^2-1}\right)} = \int_0^\infty e^{-\log\left(p+\sqrt{p^2-1}\right)u}du = \int_0^\infty \frac{1}{\left(p+\sqrt{p^2-1}\right)^u}du.$$

It is known that

$$\frac{1}{\left(p+\sqrt{p^2-1}\right)^u} = \mathcal{L}\left[\frac{u}{t}I_u(t)\right], \quad u>0.$$

Consequently

$$\frac{1}{\log\left(p+\sqrt{p^2-1}\right)} = \int_0^\infty \mathcal{L}\left[\frac{u}{t}I_u(t)\right]du = \mathcal{L}\left[\frac{1}{t}\int_0^\infty uI_u(t)du\right].$$

Applying the Faltung theorem of the Laplace transformation we obtain

$$f(t)=\frac{1}{t}\int_0^\infty uI_u(t)du-\frac{1}{t}\int_0^\infty uI_u(t)du*\int_0^t\frac{I_1(u)}{u}du, \tag{21}$$

where $*$ denotes the convolution. Taking into account (18) and

$$\frac{I_1(t)}{t}*\frac{I_n(t)}{t}=\frac{n+1}{n}\frac{I_{n+1}(t)}{t}$$

and introducing the notation

$$g(t)=\frac{\frac{e^{-t}}{\pi^2}-2\int_0^\infty\frac{u}{(u^2+\pi^2)^2}e^{-t\,\mathrm{ch}\,u}du}{t} \tag{22}$$

we obtain

$$\begin{aligned}f(t)&=g(t)+\frac{2}{\pi t}\sum_{n=1}^\infty nS_i(\pi n)I_n(t)-\\&-g(t)*\int_0^t\frac{I_1(u)}{u}du-\frac{2}{\pi}\sum_{n=1}^\infty(n+1)S_i(\pi n)\int_0^t\frac{I_{n+1}(u)}{u}du\end{aligned} \tag{23}$$

having an integrable singularity for $t=0$.

REFERENCES

[1] Cooke, J. C., Note on some integrals of Bessel functions with respect to their order, *Monatsh. Math.* **58** (1954), 1–4. *MR* **15**–792

[2] Watson, G. N., *A treatise on the theory of Bessel functions*, Cambridge University Press, Cambridge, 1944. *MR* 6–64

[3] Luke, Y. L., *Integrals of Bessel functions*, McGraw-Hill, New York – Toronto – London, 1962. *MR* **25** #5198

[4] *Handbook of mathematical functions, with formulas, graphs, and mathematical tables*, Edited by Milton Abramowitz and Irene A. Stegun, Dover Publications, Inc., New York, 1966. *MR* **34** #8606

[5] Rühs, F., Operatorenrechnung und Hadamardscher Partie finie, *Math. Nachr.* **30** (1965), 237–250. *MR* **32** #8072

[6] Schaar, G., Zur Lösung von Faltungsintegralgleichungen mit Hadamard-Integralen mittels der Mikusińskischen Operatorenrechnung, *Beiträge Anal.* **2** (1971), 99–122. *MR* **46** #9670

(Received April 9, 1992)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 197–204

ON THE FOURIER TRANSFORM OF THE BESSEL FUNCTION WITH RESPECT TO THE ORDER

T. FÉNYES

In [1] Cooke evaluated the integral

$$(1) \qquad \int_0^\infty I_\xi(z)\cos\phi\xi\, d\xi,$$

on the other hand Fényes [2] did the same for the integral

$$(2) \qquad \int_0^\infty I_\xi(z)\sin\phi\xi\, d\xi,$$

and showed that

$$(3) \qquad \int_0^\infty \xi I_\xi(t)d\xi = \frac{e^{-t}}{\pi^2} - 2\int_0^\infty \frac{u}{(u^2+\pi^2)^2} e^{-t\,\mathrm{ch}\,u}du + \frac{2}{\pi}\sum_{n=1}^\infty nS_i(\pi n)I_n(t), \quad t \geqq 0$$

where I_ξ is the modified Bessel function of the first kind:

$$S_i(x) = \int_0^x \frac{\sin u}{u}du.$$

In this paper we calculate the integrals

$$(4) \qquad \int_0^\infty J_\xi(t)\cos\phi\xi\, d\xi, \quad \int_0^\infty J_\xi(t)\sin\phi\xi\, d\xi \quad \text{and} \quad \int_0^\infty \xi J_\xi(t)d\xi, \quad t \geqq 0$$

resp., where J_ξ is the Bessel function of the first kind.

We start with the known formulas

1980 *Mathematics Subject Classification.* Primary 33C10.
Key words and phrases. Bessel functions, Fourier transform.

[1]Research (partially) supported by the National Foundation for Scientific Research Grant No. 6032/6319.

$$J_\xi(t)=\frac{1}{\pi}\int_0^\pi \cos(\xi\Theta - t\sin\Theta)d\Theta - \frac{\sin\xi\pi}{\pi}\int_0^\infty e^{-t\,\mathrm{sh}\,u-\xi u}du, \tag{5}$$

$$\begin{aligned}\cos(t\sin\Theta) &= J_0(t)+2\sum_{k=1}^{\infty}J_{2k}(t)\cos 2k\Theta,\\ \sin(t\sin\Theta) &= 2\sum_{k=0}^{\infty}J_{2k+1}(t)\sin(2k+1)\Theta\end{aligned} \tag{6}$$

(see Watson [3], Luke [4], Abramowitz–Stegun [5]). So the first integral in [5] can be written as

$$\begin{aligned}&\frac{1}{\pi}\int_0^\pi \cos\xi\Theta\Big(J_0(t)+2\sum_{k=1}^{\infty}J_{2k}(t)\cos 2k\Theta\Big)d\Theta+\\ &+\frac{2}{\pi}\int_0^\pi \sin\xi\Theta\sum_{k=0}^{\infty}J_{2k+1}(t)\sin(2k+1)\Theta d\Theta.\end{aligned} \tag{7}$$

Applying elementary trigonometrical additional formulas, by a legitimate termwise integration of the infinite series, (7) can be reduced to the form

$$\begin{aligned}&\frac{1}{\pi}\frac{\sin\xi\pi}{\xi}J_0(t)+\frac{\sin\xi\pi}{\pi}\sum_{k=1}^{\infty}J_{2k}(t)\Big(\frac{1}{\xi+2k}+\frac{1}{\xi-2k}\Big)+\\ &+\frac{\sin\xi\pi}{\pi}\sum_{k=0}^{\infty}J_{2k+1}(t)\Big(\frac{1}{\xi+2k+1}-\frac{1}{\xi-(2k+1)}\Big).\end{aligned} \tag{8}$$

We write

$$\int_0^\infty J_\xi(t)\cos\phi\xi d\xi = g_1(\phi)+g_2(\phi), \tag{9}$$

where

$$\begin{aligned}g_1(\phi)=\frac{J_0(t)}{\pi}\int_0^\infty \frac{\sin\pi\xi\cos\phi\xi}{\xi}d\xi+\frac{1}{\pi}\int_0^\infty \sin\pi\xi\cos\phi\xi\sum_{k=1}^{\infty}J_{2k}(t)\times\\ \times\Big(\frac{1}{\xi+2k}+\frac{1}{\xi-2k}\Big)d\xi+\\ +\frac{1}{\pi}\int_0^\infty \sin\pi\xi\cos\phi\xi\sum_{k=0}^{\infty}J_{2k+1}(t)\Big(\frac{1}{\xi+2k+1}-\frac{1}{\xi-(2k+1)}\Big)d\xi,\end{aligned} \tag{10}$$

$$g_2(\phi) = -\frac{1}{\pi}\int_0^\infty \sin \pi\xi \cos \phi\xi \int_0^\infty e^{-t \operatorname{sh} u - \xi u} du\, d\xi. \tag{11}$$

By the inversion of the integration in (11) we obtain

$$g_2(\phi) = -\frac{1}{2\pi}\int_0^\infty \Big(\frac{\pi+\phi}{(\pi+\phi)^2+u^2} + \frac{\pi-\phi}{(\pi-\phi)^2+u^2}\Big) e^{-t \operatorname{sh} u} du. \tag{12}$$

The value of the first term of $g_1(\phi)$ is

$$\begin{cases} \dfrac{J_0(t)}{2}, & \text{for} \quad |\phi| < \pi, \\ 0, & \text{for} \quad |\phi| > \pi, \\ \dfrac{J_0(t)}{4}, & \text{for} \quad |\phi| = \pi. \end{cases} \tag{13}$$

Let $|\phi| \neq \pi$. In (10) we may interchange the order of integration and summation, so we obtain the expression

$$\begin{aligned} &\frac{1}{\pi}\sum_{k=1}^{\infty} J_{2k}(t) \int_0^\infty \Big(\frac{\sin \pi\xi \cos \phi\xi}{\xi+2k} + \frac{\sin \pi\xi \cos \phi\xi}{\xi-2k}\Big) d\xi + \\ &+\frac{1}{\pi}\sum_{k=0}^{\infty} J_{2k+1}(t) \int_0^\infty \Big(\frac{\sin \pi\xi \cos \phi\xi}{\xi+2k+1} - \frac{\sin \pi\xi \cos \phi\xi}{\xi-(2k+1)}\Big) d\xi. \end{aligned} \tag{14}$$

We should evaluate the integrals

$$\begin{aligned} I_{2k} &= \int_0^\infty \sin \pi\xi \cos \phi\xi \Big(\frac{1}{\xi+2k} + \frac{1}{\xi-2k}\Big) d\xi, \\ I_{2k+1} &= \int_0^\infty \sin \pi\xi \cos \phi\xi \Big(\frac{1}{\xi+2k+1} - \frac{1}{\xi-(2k+1)}\Big) d\xi. \end{aligned} \tag{15}$$

Applying again elementary trigonometrical additional formulas, and introducing the substitutions $\xi + 2k = u$, $\xi - 2k = u$, $\xi + 2k + 1 = u$, and $\xi - (2k + 1) = u$, respectively, after some routine steps (15) can be reduced to the form

$$I_{2k}=\frac{\cos 2k\phi}{2}\left[\int_{2k}^{\infty}\frac{\sin(\pi+\phi)u+\sin(\pi-\phi)u}{u}du+\int_{-2k}^{\infty}\frac{\sin(\pi+\phi)u+\sin(\pi-\phi)u}{u}du\right]+$$

(16)

$$+\frac{\sin 2k\phi}{2}\left[\int_{2k}^{\infty}\frac{\cos(\pi-\phi)u-\cos(\pi+\phi)u}{u}du-\int_{-2k}^{\infty}\frac{\cos(\pi-\phi)u-\cos(\pi+\phi)u}{u}du\right],$$

$$I_{2k+1}=-\frac{\cos(2k+1)\phi}{2}\left[\int_{2k+1}^{\infty}\frac{\sin(\pi+\phi)u+\sin(\pi-\phi)u}{u}du-\right.$$

(17)

$$\left.-\int_{-2k-1}^{\infty}\frac{\sin(\pi+\phi)u+\sin(\pi-\phi)u}{u}du\right]+\frac{\sin(2k+1)\phi}{2}\times$$

$$\times\left[\int_{2k+1}^{\infty}\frac{\cos(\pi+\phi)u-\cos(\pi-\phi)u}{u}du+\int_{-2k-1}^{\infty}\frac{\cos(\pi+\phi)u-\cos(\pi-\phi)u}{u}du\right].$$

So we get

$$(18)\quad I_{2k}=\cos 2k\phi\int_{0}^{\infty}\frac{\sin(\pi+\phi)u+\sin(\pi-\phi)u}{u}du=\begin{cases}\pi\cos 2k\phi, & \text{for } |\phi|<\pi,\\ 0, & \text{for } |\phi|>\pi.\end{cases}$$

$$I_{2k+1}=\cos(2k+1)\phi[S_i((2k+1)(\pi+\phi))+S_i((2k+1)(\pi-\phi))]+$$

$$+\sin(2k+1)\phi\left[\int_{0}^{\infty}\frac{\cos(\pi+\phi)u-\cos(\pi-\phi)u}{u}du+\right.$$

(19)

$$\left.+\int_{0}^{2k+1}\frac{\cos(\pi-\phi)u-\cos(\pi+\phi)u}{u}du\right]=$$

$$=\cos(2k+1)\phi\,[S_i((2k+1)(\pi+\phi))+S_i((2k+1)(\pi-\phi))]+$$

$$+\sin(2k+1)\phi\left[\log\left|\frac{\pi-\phi}{\pi+\phi}\right|+S_1((2k+1)(\pi+\phi))-S_1((2k+1)(\pi-\phi))\right].$$

Here is

$$S_1(x)=\int_{0}^{x}\frac{1-\cos u}{u}du.$$

Consequently, for $|\phi|<\pi$, we have by (10), (13), (14), (15), (18), (19)

$$g_1(\phi)=\frac{J_0(t)}{2}+\sum_{k=1}^{\infty}J_{2k}(t)\cos 2k\phi+\frac{1}{\pi}\log\left|\frac{\pi-\phi}{\pi+\phi}\right|\sum_{k=0}^{\infty}J_{2k+1}(t)\sin(2k+1)\phi+$$

$$(20)\quad +\frac{1}{\pi}\sum_{k=0}^{\infty}J_{2k+1}(t)\sin(2k+1)\phi\,[S_1((2k+1)(\pi+\phi))-S_1((2k+1)(\pi-\phi))]+$$

$$+\frac{1}{\pi}\sum_{k=0}^{\infty}J_{2k+1}(t)\cos(2k+1)\phi\,[S_i((2k+1)(\pi+\phi))+S_i((2k+1)(\pi-\phi))].$$

Taking (6) into account we get

$$g_1(\phi)=\frac{\cos(t\sin\phi)}{2}+\frac{1}{2\pi}\log\left|\frac{\pi-\phi}{\pi+\phi}\right|\sin(t\sin\phi)+$$

$$(21)\quad +\frac{1}{\pi}\sum_{k=0}^{\infty}J_{2k+1}(t)\sin(2k+1)\phi\,[S_1((2k+1)(\pi+\phi))-S_1((2k+1)(\pi-\phi))]+$$

$$+\frac{1}{\pi}\sum_{k=0}^{\infty}J_{2k+1}(t)\cos(2k+1)\phi\,[S_i((2k+1)(\pi+\phi))+S_i((2k+1)(\pi-\phi))].$$

By (9), (12) we have for $|\phi|<\pi$

$$\int_0^{\infty}J_{\xi}(t)\cos\phi\xi\,d\xi=\frac{\cos(t\sin\phi)}{2}+\frac{1}{2\pi}\log\left|\frac{\pi-\phi}{\pi+\phi}\right|\sin(t\sin\phi)+$$

$$+\frac{1}{\pi}\sum_{k=0}^{\infty}J_{2k+1}(t)\sin(2k+1)\phi\,[S_1((2k+1)(\pi+\phi))-S_1((2k+1)(\pi-\phi))]+$$

$$(22)\qquad +\frac{1}{\pi}\sum_{k=0}^{\infty}J_{2k+1}(t)\cos(2k+1)\phi\,[S_i((2k+1)(\pi+\phi))+S_i((2k+1)(\pi-\phi))]-$$

$$-\frac{1}{2\pi}\int_0^{\infty}\left(\frac{\pi+\phi}{(\pi+\phi)^2+u^2}+\frac{\pi-\phi}{(\pi-\phi)^2+u^2}\right)e^{-t\,\mathrm{sh}\,u}\,du.$$

It is easily seen that for $|\phi|>\pi$ the first term is replaced by zero. Since

$$\log\left|\frac{\pi-\phi}{\pi+\phi}\right|\sin(t\sin\phi)$$

tends to zero for $|\phi|\to\pi$, applying Fourier's theorem, we get that by replacing the first term by $\frac{1}{4}$, (22) also holds for $|\phi|=\pi$. The above infinite series

converge uniformly with respect to ϕ, which can easily be seen by a simple majorization.

In particular for $\phi = 0$

$$\int_0^\infty J_\xi(t)d\xi = \frac{1}{2} + \frac{2}{\pi}\sum_{k=0}^{\infty} J_{2k+1}(t)S_i[(2k+1)\pi] - \int_0^\infty \frac{e^{-t\,\mathrm{sh}\,u}}{\pi^2+u^2}du. \tag{23}$$

Let us consider the second integral in [4]. Applying [5], [8] we can write

$$\int_0^\infty J_\xi(t)\sin\phi\xi\, d\xi = h_1(\phi) + h_2(\phi), \tag{24}$$

where, for $|\phi| \neq \pi$,

$$\begin{aligned} h_1(\phi) = \frac{J_0(t)}{\pi}\int_0^\infty \frac{\sin\pi\xi\sin\xi\phi}{\xi}d\xi + \frac{1}{\pi}\int_0^\infty \sin\pi\xi\sin\phi\xi\times \\ \times\Big\{\sum_{k=1}^{\infty} J_{2k}(t)\Big(\frac{1}{\xi+2k}+\frac{1}{\xi-2k}\Big)d\xi + \\ +\sum_{k=0}^{\infty} J_{2k+1}(t)\Big(\frac{1}{\xi+2k+1}-\frac{1}{\xi-(2k+1)}\Big)d\xi\Big\}, \end{aligned} \tag{25}$$

$$h_2(\phi) = -\frac{1}{\pi}\int_0^\infty \sin\pi\xi\sin\phi\xi\int_0^\infty e^{-t\,\mathrm{sh}\,u-\xi u}du\,d\xi. \tag{26}$$

Though $h_1(\phi)$, $h_2(\phi)$ have singularities if $\phi = \pm\pi$, we show that their sum is continuous for $\phi = \pm\pi$. The evaluation of $h_1(\phi)$ is wholly analogous to that of $g_1(\phi)$ so we omit here the details. We obtain the result:

If $|\phi| < \pi$, then

$$\begin{aligned} h_1(\phi) = \frac{\sin(t\sin\phi)}{2} + \frac{1}{2\pi}\log\Big|\frac{\pi+\phi}{\pi-\phi}\Big|\cos(t\sin\phi) + \\ +\frac{1}{\pi}\sum_{k=1}^{\infty} J_{2k}(t)\sin 2k\phi(S_i[2k(\pi+\phi)] + S_i[2k(\pi-\phi)]) + \\ +\frac{1}{\pi}\sum_{k=1}^{\infty} J_{2k}(t)\cos 2k\phi(S_1[2k(\pi-\phi)] - S_1[2k(\pi+\phi)]), \end{aligned} \tag{27}$$

and for $|\phi| > \pi$ the first term in (27) is replaced by zero. By interchanging the order of integration in (26) we have

$$h_2(\phi) = -\frac{1}{2\pi}\int_0^\infty \left(\frac{u}{u^2+(\pi-\phi)^2} - \frac{u}{u^2+(\phi+\pi)^2}\right) e^{-t\,\mathrm{sh}\,u}du, \tag{28}$$

and an integration by parts gives

(29)

$$h_2(\phi) = -\frac{1}{2\pi}\log\left|\frac{\pi+\phi}{\pi-\phi}\right| - \frac{t}{4\pi}\int_0^\infty \log\frac{u^2+(\phi-\pi)^2}{u^2+(\phi+\pi)^2}\,\mathrm{ch}\,u\,e^{-t\,\mathrm{sh}\,u}du, \quad (|\phi| \neq \pi).$$

Here the integral does not exist for $t = 0$, however, the whole second term tends to zero if $t \to 0$. Consequently, by (24), (27), (29)

$$\begin{aligned}\int_0^\infty J_\xi(t)\sin\phi\xi\,d\xi = {} & \frac{\sin(t\sin\phi)}{2} + \frac{1}{2\pi}\log\left|\frac{\pi+\phi}{\pi-\phi}\right|(\cos(t\sin\phi)-1)+ \\ & +\frac{1}{\pi}\sum_{k=1}^\infty J_{2k}(t)\sin 2k\phi(S_i[2k(\pi+\phi)]+S_i[2k(\pi-\phi)])+ \\ & +\frac{1}{\pi}\sum_{k=1}^\infty J_{2k}(t)\cos 2k\Phi(S_1[2k(\pi-\phi)]-S_1[2k(\pi+\phi)])- \\ & -\frac{t}{4\pi}\int_0^\infty \log\frac{u^2+(\Phi-\pi)^2}{u^2+(\phi+\pi)^2}\,\mathrm{ch}\,u\,e^{-t\,\mathrm{sh}\,u}du, \quad \text{for} \quad |\phi|<\pi,\end{aligned} \tag{30}$$

and $\frac{\sin(t\sin\phi)}{2}$ is replaced by zero in (30) if $|\phi| > \pi$.

Since

$$\log\left|\frac{\pi+\phi}{\pi-\phi}\right|(\cos(t\sin\phi)-1)$$

tends to zero for $|\phi| \to \pi$, applying Fourier's theorem it can easily be seen that (30) also holds for $|\phi| = \pi$.

Let us now consider the integral

$$\int_0^\infty \xi\cos\phi\xi J_\xi(t)d\xi,$$

which obviously converges uniformly since

$$\left|\int_0^\infty \xi \cos\phi\xi J_\xi(t)d\xi\right| \leqq \int_0^\infty \xi|J_\xi(t)|d\xi \leqq \int_0^\infty \xi I_\xi(t)d\xi.$$

So by an elementary property of the Fourier transformation with the aid of (30) we obtain

$$\int_0^\infty \xi J_\xi(t)d\xi = \left[\frac{d}{d\phi}\int_0^\infty J_\xi(t)\sin\phi\xi\, d\xi\right]_{\phi=0} = \tag{31}$$

$$= \frac{t}{2} + \frac{4}{\pi}\sum_{k=1}^\infty kJ_{2k}(t)S_i(2k\pi) + t\int_0^\infty \frac{\operatorname{ch} u\, e^{-t\operatorname{sh} u}du}{u^2+\pi^2},$$

and by an integration by parts we obtain the final result

$$\int_0^\infty \xi J_\xi(t)d\xi = \frac{t}{2} + \frac{1}{\pi^2} + \frac{4}{\pi}\sum_{k=1}^\infty kJ_{2k}(t)S_i(2k\pi) - 2\int_0^\infty \frac{u}{(u^2+\pi^2)^2}e^{-t\operatorname{sh} u}du. \tag{32}$$

Here the differentiation under the integral sign is legitimate for every $t>0$ and so is the termwise differentiation of the infinite series in (30) since the obtained integral and infinite series are uniformly convergent, as it can easily be seen by a simple majorization.

(31), (32) are very convenient for numerical calculations. It is easy to obtain numerical values for the integrals for any fixed t. On the other hand, the functions $J_{2k}(t)$, $J_{2k+1}(t)$, $S_i[2k\pi]$, $S_i[(2k+1)\pi]$ are tabulated (see Abramowitz–Stegun [5], Luke [4]).

REFERENCES

[1] Cooke, J. C., Note on some integrals of Bessel functions with respect to their order, *Monatsh. Math.* **58** (1954), 1–4. *MR* **15**–792

[2] Fényes, T., On the Fourier transform of the modified Bessel function with respect to the order, *Studia Sci. Math. Hungar.* **28** (1993), 189–196.

[3] Watson, G. N., *A treatise on the theory of Bessel functions*, Cambridge University Press, Cambridge, 1944. *MR* **6**–64

[4] Luke, Y. L., *Integrals of Bessel functions*, McGraw-Hill, New York – Toronto – London, 1962. *MR* **25** #5198

[5] *Handbook of mathematical functions, with formulas, graphs, and mathematical tables*, Edited by Milton Abramowitz and Irene A. Stegun, Dover Publications, Inc., New York, 1966. *MR* **34** #8606

(Received May 26, 1992)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 205–213

TRUNCATED HERMITE INTERPOLATION POLYNOMIALS

P. VÉRTESI* and Y. XU

Abstract

The convergence behaviour of the truncated Hermite interpolation polynomials is discussed in both the weighted L^p norm and the uniform norm. The rate of convergence is given in terms of the modulus $\omega_\varphi(f;t)$ of Ditzian and Totik.

1. Introduction

Let w be a Jacobi weight function defined by $w(x)=(1-x)^\alpha(1+x)^\beta$, $|x|\leqq 1$, $\alpha>-1$, $\beta>-1$, and $w(x)=0$, $|x|\geqq 1$. Let $p_n(w,x)$ be the Jacobi polynomials orthonormal with respect to w. Let $x_{kn}(w)$ be the zeros of $p_n(w,x)$ with

$$-1<x_{nn}(w)<x_{n,n-1}(w)<\ldots<x_{n,1}(w)<1. \tag{1.1}$$

For any given integer $m\geqq 1$, let $H_{mn}(w,f)$ be the Hermite interpolating operator, which is defined to be the unique polynomial of degree at most $mn-1$ satisfying

$$H_{mn}^{(j)}(w,f,x_{kn})=f^{(j)}(x_{kn}),\quad 1\leqq k\leqq n,\ 0\leqq j\leqq m-1, \tag{1.2}$$

where $x_{kn}=x_{kn}(w)$.

The mean convergence of this Hermite interpolating polynomial has been studied recently by the authors [13]. However, the uniform convergence has not been fully studied for an arbitrary given integer m. The only result we know is due to Esser and Scherer [2], which states that for the zeros of Chebyshev polynomial, the rate of convergence of $H_{mn}(w,f)$ for $f\in C^{m-1}$ is $(\log n/n)^{m-1}\omega_p(f^{(m-1)};n^{-1})$, where $\omega_p(f,t)$ is the usual pth modulus of continuity. In the investigating of the mean convergence of $H_{mn}(w,f)$ [13], we realized that our approach is really like the uniform method. It is the

1991 *Mathematics Subject Classifications.* Primary 41A05; Secondary 41A10, 41A25.
Key words and phrases. Interpolation, Jacobi polynomials, mean convergence.

*Research supported by the Hungarian National Foundation for Scientific Research Grant Nos. 1801, 1910 and T7570.

purpose of this paper to unify the mean convergence and the uniform convergence method, to show the similarity between the two cases. In doing this, we shall consider a more general situation, namely, the interpolating polynomial $H_{r,m,n}(w,f)$ which is defined to be the unique polynomial of degree $mn-1$ satisfying the following conditions

$$\begin{aligned} H^{(j)}_{r,m,n}(w,f,x_{kn}) &= f^{(j)}(x_{kn}), \quad 0 \leqq j \leqq r, \\ H^{(j)}_{r,m,n}(w,f,x_{kn}) &= 0, \quad r < j \leqq m-1, \end{aligned} \tag{1.3}$$

where r is a fixed integer, $0 \leqq r \leqq m-1$. When $r = m-1$, we drop the second line of (1.3), and write $H_{m-1,m,n}(w,f) = H_{mn}(w,f)$. We shall call $H_{r,m,n}(w,f)$ the *truncated Hermite interpolation polynomials*, while $r=0$, $H_{0,m,n}(w,f)$ is the Hermite–Fejér interpolation of higher order.

We shall present mean convergence results for all $m \geqq 2$, $r \geqq 0$ and uniform convergence results for all $m \geqq 2$, $r > 0$. The case $r = 0$ in the uniform convergence is of some different character, it has been discussed by Vértesi [10] and Xu [14]. We shall prove only sufficient conditions on the rate of convergence, and try to concentrate more on ideas and methods. Let us note here that the previous results [10], [11] and [12] are mostly "if and only if" statements. For mean convergence when $m=1$ and $m=2$, see [4, 5].

The results of this paper generalize those in [7], [8], [11] and [12]. Throughout this paper we shall apply the $\omega_\varphi(f;t)$ modulus (see [1]) which seems to be more natural than the usual one (cf. [14]). Using ω_φ, even in the previously proved special cases, we have better results.

2. Main results

Throughout this paper, we shall adopt the following convention. The letters $c, c_1, \ldots$ will denote positive constants, being independent of variables and indices, unless otherwise indicated. Their value may be different at different occurrences, even in subsequent formulae. We define L^p $(0 < p < +\infty)$ in the usual way: spaces of real variable functions on $[-1,1]$. For sake of convenience we use the norm notation $\|\cdot\|_p$ even if $0 < p < 1$. $\|\cdot\|_\infty$ denotes the maximum norm on $[-1,1]$, while

$$\|f\|_{u,p} := \left(\int_{-1}^{1} |f(x)u(x)|^p \, dx \right)^{1/p}. \tag{2.1}$$

Furthermore, $\|f^{(j)}\|_*$ is defined by

$$\|f^{(j)}\|_* := \max_{1 \leqq k \leqq n} \left| \left(\sqrt{1-x_{kn}^2} \right)^j f^{(j)}(x_{kn}) \right|, \tag{2.2}$$

for $0 \leqq j \leqq m-1$, where m is a fixed integer. Let $\varphi(x) = \sqrt{1-x^2}$. The weighted modulus of continuity of f defined by Ditzian and Totik [1] is given by

$$\omega_\varphi(f;t) = \sup_{0<h\leqq t} \|f(\cdot + h\varphi(\cdot)/2) - f(\cdot - h\varphi(\cdot)/2)\|_\infty,$$

where if $x \pm h\varphi(x)/2 \notin (-1,1)$, the expression inside $\|\cdot\|$ is taken to be zero. $E_n(f)$ is defined by

$$E_n(f) = \min \|f - P\|_\infty$$

where the minimum is taken over all polynomials of degree at most n.

As before, w denotes the weight function with which the interpolating knots associated, α and β will always be the parameters of w. We define

$$\Gamma = \max\{\alpha,\beta\}, \quad \gamma = \min\{\alpha,\beta\}, \tag{2.3}$$

and

$$A_m = -\frac{1}{2} - \frac{2}{m}, \quad C_m = -\frac{1}{2} - \frac{1}{m}. \tag{2.4}$$

The following assumptions on α and β will be used throughout the paper (see [10, 11] for the reasoning)

$$\gamma \geqq A_m, \quad \text{for odd integer} \quad m, \tag{2.5a}$$

and

$$\gamma \geqq C_m \text{ or } A_m \leqq \gamma < C_m \text{ and } \Gamma - \gamma \leqq \frac{2}{m}, \text{ for even integer } m. \tag{2.5b}$$

When we deal with uniform norm, we assume more, namely,

$$\gamma > A_m, \text{ for odd integer } m. \tag{2.6}$$

Finally, we define

$$w_m(x) = \left(w(x)\sqrt{1-x^2}\right)^{m/2}.$$

We now state our main results.

THEOREM 2.1. *Let $m \geqq 1$, $0 \leqq r \leqq m-2$. Let $0 < p < \infty$, u be a Jacobi weight, $u \in L^p$. Then, for every integer $\lambda \geqq 0$, any $f \in C^r$,*

$$\text{(i)} \quad \|H_{r,m,n}(w,f) - f\|_\infty \leqq \frac{c}{n^{r-\lambda}}\left(f^{(r)};\frac{1}{n}\right)\left(1 + \frac{\log n}{n^\lambda}\right), \tag{2.7}$$

if

$$w_m^{-1}\varphi^\lambda \in L^\infty; \tag{2.8}$$

$$\text{(ii)} \quad \|H_{r,m,n}(w,f) - f\|_\infty \leqq \frac{c}{n^{r-\lambda}}\omega_\varphi\left(f^{(r)};\frac{1}{n}\right), \tag{2.9}$$

if $\gamma > C_m$ *and*

$$w_m^{-1}\varphi^\lambda u^{1/p} \in L^p. \tag{2.10}$$

THEOREM 2.2. *Let* $m \geqq 1$. *Let* $0 < p < +\infty$, u *be a Jacobi weight,* $u \in L^p$. *Then, for every integer* $\lambda \geqq 0$, *any* $f \in C^{m-1}$,

$$\text{(i)} \quad \|H_{m,n}(w,f) - f\|_\infty \leqq \frac{c}{n^{m-1-\lambda}} E_n(f^{(m-1)})\Big(1 + \frac{\log n}{n^\lambda}\Big), \tag{2.11}$$

if

$$w_m^{-1}\varphi^\lambda \in L^\infty; \tag{2.12}$$

$$\text{(ii)} \quad \|H_{m,n}(w,f) - f\|_{u,p} \leqq \frac{c}{n^{m-1-\lambda}} E_n(f^{(m-1)}), \tag{2.13}$$

if $\gamma > C_m$ *and*

$$w_m^{-1}\varphi^\lambda u^{1/p} \in L^p. \tag{2.14}$$

By choosing $\lambda = r - 1$ in the above theorems, we have convergence for the largest range of α and β (cf. (2.8) and (2.10)). By choosing $\lambda = 0$, we have the fastest rate of convergence. Theorem 2.2 with $\lambda = m - 1$ improves the result of [2] mentioned before. Further remarks and results are given in Section 4.

3. Proofs

The Hermite interpolating polynomial (1.2) can be written as

$$H_{mn}(w,f,x) = \sum_{t=0}^{m-1} \sum_{k=1}^{n} f^{(t)}(x_{kn}) h_{tkm}(x), \quad m = 1,2,\ldots \tag{3.1}$$

where h_{tkm} are the polynomials of degree $mn-1$, which are uniquely defined by

$$h_{tkm}^{(i)}(x_{jn}) = \delta_{ti}\delta_{tj}, \quad 0 \leqq i, t \leqq m-1,\ 1 \leqq j\, k \leqq n. \tag{3.2}$$

From (2.1) we have that $H_{r,m,n}(w,f)$ defined in (1.3) can be written as

$$H_{r,m,n}(w,f,x) = \sum_{t=0}^{r} \sum_{k=1}^{n} f^{(t)}(x_{kn}) h_{tkm}(x). \tag{3.3}$$

Let j be the index determined by $|x_j - x| = \min\limits_{1 \leqq k \leqq n} |x - x_{kn}|$. Our main lemmas are the following

LEMMA 3.1. *Let $t \geqq 0$, $m-t$ be odd integers, $0 \leqq r \leqq m-1$. If $t > r$, $\gamma \geqq A_m$ or $t = r$, $\gamma > A_m$, then*

$$\sum_{k=1}^{n} \left(\frac{n}{\varphi(x_{kn})}\right)^{t-r} |h_{tkm}(x)| \leqq \frac{c}{n^r}\left[(\varphi(x_{jn}))^r \log n + w_m(x_{jn})^{-1}\right], \tag{3.4}$$

and if $t = r$, $\gamma = A_m$, then

$$\sum_{k=1}^{n} |h_{rkm}(x)| \leqq c\frac{\log n}{n^r}\left[\varphi(x_{jn})^r + w_m(x_{jn})^{-1}\right]. \tag{3.5}$$

LEMMA 3.2. *Let $t \geqq 0$, $m-t$ be even integers, $0 \leqq r \leqq m-1$. Then for $t \geqq r$,*

$$\sum_{k=1}^{n} \left(\frac{n}{\varphi(x_{kn})}\right)^{t-r} |h_{tkm}(x)| \leqq \frac{c}{n^r}\left[\varphi(x_{jn})^r + \frac{\log n}{n} w_m(x_{mn})^{-1}\right]. \tag{3.6}$$

PROOF. The proof of these two lemmas can be reduced to the proof of [11, Lemma 3.2]. Actually, if along the lines of the proof in [11], we write down the first one or two estimates on these quantities, we will see that (3.4) follows from the estimates there for $t = r$. So does (3.5). The case $r = 0$ has been considered in [13], too.

We also need the following lemma proved recently by the second author [15].

LEMMA 3.3. *For $r \geqq 0$, $f \in C^r$, there exists a polynomial P_n of degree at most n such that for all $-1 \leqq x \leqq 1$,*

$$|f^{(t)}(x) - P_n^{(t)}(x)| \leqq c[\Delta_n(x)]^{r-t} E_{n-r}(f^{(r)}), \quad 0 \leqq t \leqq r \tag{3.7}$$

and

$$|P_n^{(t)}(x)| \leqq c[\Delta_n(x)]^{r-t} \omega_\varphi\left(f^{(r)}; \frac{1}{n}\right), \quad t > r, \tag{3.8}$$

where $\Delta_n(x) = \varphi(x)n^{-1} + n^{-2}$.

PROOF OF THEOREM 2.1. We first prove the uniform estimate (2.7). By (3.3), we have from Lemmas 3.1 and 3.2 with $r = 0$ that

$$|H_{r,m,n}(w, f, x)| \leqq \sum_{t=0}^{r} \|f^{(t)}\|_* \sum_{k=1}^{n} \left(\frac{1}{\varphi(x_{kn})}\right)^t |h_{tk}(x)| \leqq$$

$$\leqq cn^\lambda \sum_{t=0}^{r} \|f^{(t)}\|_* \frac{1}{n^t}\left(\frac{\log n}{n^\lambda} + w_m(x_{jn})^{-1}\varphi(x_{jn})^\lambda\right)$$

where we use $n\varphi(x_{jn}) \geqq c$ which follows from

$$\theta_{k+1,n} - \theta_{kn} \sim \frac{1}{n}$$

with $x_{kn} = \cos\theta_{kn}$, $0 \leqq k \leqq n+1$, $x_{0n} = 1$, $x_{n+1,n} = -1$ (cf. [5]). Therefore under condition (2.8) we have

$$\|H_{r,m,n}(w,f)\|_\infty \leqq cn^\lambda \sum_{t=0}^{r} \|f^{(t)}\|_* \frac{1}{n^t}\left(1 + \frac{\log n}{n^\lambda}\right). \tag{3.9}$$

Now let P_n be the polynomial in Lemma 3.3. By the triangular inequality, we have

(3.10)

$$\|H_{r,m,n}(w,f)-f\|_\infty \leqq \|H_{r,m,n}(w,f-P_n)\|_\infty + \|H_{r,m,n}(w,P_n)-P_n\|_\infty + \|P-f\|_\infty.$$

From (3.7) and (3.9) it follows that

$$\|H_{r,m,n}(w,f-P_n)\|_\infty \leqq cE_n(f^{(r)})\frac{1}{n^{r-\lambda}}\left(1+\frac{\log n}{n^\lambda}\right).$$

Since

$$E_n(f) \leqq \frac{c}{n^r}E_n(f^{(r)}) \leqq \frac{c}{n^r}\omega_\varphi\left(f^{(r)};\frac{1}{n}\right)$$

where the last inequality follows from [1, Theorem 7.1.1], we only need to estimate the second term in (3.10). Since $H_{mn}(w,f)$ preserves polynomial of degree $mn-1$, we have from (3.1) and (3.3) that

$$P_n(x) - H_{r,m,n}(w,P_n,x) = \sum_{t=r+1}^{m-1}\sum_{k=1}^{n} P_n^{(t)}(x_{kn})h_{tkm}(x).$$

Therefore, by (3.8), Lemmas 3.1 and 3.2,

$$|P_n(x) - H_{r,m,n}(w,P_n,x)| \leqq$$

$$\leqq \sum_{t=r+1}^{m-1}\frac{1}{n^{t-r}}\sum_{k=1}^{n}\left|\varphi(x_{kn})^{t-r}P_n^{(t)}(x_{kn})\right|\left(\frac{n}{\varphi(x_{kn})}\right)^{t-r}|h_{tkm}(x)| \leqq$$

$$\leqq \frac{c}{n^r}\omega_\varphi\left(f^{(r)};\frac{1}{n}\right)\left[\varphi(x_{jn})^r\log n + w_m(x_{jn})^{-1}\right] \leqq$$

$$\frac{c}{n^{r-\lambda}}\omega_\varphi\left(f^{(r)};\frac{1}{n}\right)\left(1+\frac{\log n}{n^\lambda}\right)$$

where in the last inequality, we use $n\varphi(x_{jn}) \geqq c$ and condition (2.8) again. Thus, the proof of (2.7) is completed.

We now prove the mean convergence part (2.9). In [13] we proved that under condition (2.10) and $\gamma > C_m$,

$$\|H_{nm}(w,f)\|_{u,p} \leqq cn^{\lambda} \sum_{t=0}^{m-1} \frac{1}{n^t} \|f^{(t)}\|_*,$$

[13, Lemma 4.2]. Since $H_{mn}(w,f)$ preserves polynomial of degree $mn-1$, we have that for any polynomial p of degree $\leqq mn-1$,

$$\|p\|_{u,p} \leqq cn^{\lambda} \sum_{t=0}^{m-1} \frac{1}{n^t} \|p^{(t)}\|_* \tag{3.11}$$

under condition (2.10) and $\gamma > C_m$. By (1.3) and (3.11) with $p = H_{r,m,n}(w,f)$, we then have

$$\|H_{r,m,n}(w,f)\|_{u,p} \leqq cn^{\lambda} \sum_{t=0}^{r} \frac{1}{n^t} \|f^{(t)}\|_*. \tag{3.12}$$

Let P be the polynomial in Lemma 3.3. From (3.7) and (3.12) we have

$$\|H_{r,m,n}(w,f-P)\|_{u,p} \leqq cn^{\lambda} \sum_{t=0}^{r} \frac{1}{n^t} \|f^{(t)} - P^{(t)}\|_* \leqq \frac{c}{n^{r-\lambda}} E_n(f^{(r)}).$$

Using (3.8) and (3.11) with $p = P - H_{rmn}(w,P)$, we have from (1.3) that

$$\|P - H_{rmn}(w,P)\|_{u,p} \leqq cn^{\lambda} \sum_{t=r+1}^{m-1} \frac{1}{n^t} \|P^{(t)}\|_* \leqq \frac{c}{n^{r-\lambda}} \omega_{\varphi}\left(f^{(r)}; \frac{1}{n}\right).$$

Therefore the desired result follows from the triangular inequality

$$\begin{aligned} \|H_{r,m,n}(w,f) - f\|_{u,p} \leqq \|H_{r,m,n}(w,f-P)\|_{u,p} + \\ + \|P - H_{r,m,n}(w,P)\|_{u,p} + \|f - P\|_{u,p} \end{aligned}$$

and by $E_n(f^{(r)}) \leqq c\omega_{\varphi}(f^{(r)}; 1/n)$.

PROOF OF THEOREM 2.2. The proof is similar to that of Theorem 2.1 but simpler, $H_{mn}(w,f)$ preserves polynomial, we do not need to consider the term $P - H_{nm}(w,P)$.

4. Further results and remarks

If we apply Bernstein–Markoff inequality for L^p, $0 < p < +\infty$ (cf. [6]), we have from Theorems 2.1 and 2.2 the following

THEOREM 4.1. *Let $m \geqq 1$, $\gamma > C_m$ and $0 < p < +\infty$. Let u be a Jacobi weight, $u\varphi^{-jp} \in L^1$, for a fixed j, $0 \leqq j \leqq r$. Then for each integer $\lambda \geqq 0$,*

$$\|H^{(j)}_{r,m,n}(w,f) - f^{(j)}\|_{u,p} \leqq \frac{c}{n^{r-\lambda-j}}\omega_\varphi\Big(f^{(r)};\frac{1}{n}\Big), \quad \forall f \in C^r,\ 0 \leqq r \leqq m-2,$$

and

$$\|H^{(j)}_{mn}(w,f) - f^{(j)}\|_{u,p} \leqq \frac{c}{n^{m-\lambda-j-1}}E_n(f^{(m-1)}), \quad \forall f \in C^{m-1},$$

if

$$w_m^{-1}\varphi^{\lambda-j}u^{1/p} \in L^p.$$

For the counterpart of the uniform norm, we must use $\|\varphi^j(H^{(j)}_{r,m,n}(w,f) - f^{(j)})\|_\infty$, and the reader can state the results quite easily from Theorem 2.1 by using Bernstein–Markoff inequality. For the norm of derivatives without weights, see [16] when $m = 2$, which partially justifies why we do not have an analogue of Theorem 4.1 under the unweighted norm.

THEOREM 4.2. *Let $m \geqq 1$ be a given integer. Then*

$$\|w_m(H_{r,m,n}(w,f) - f)\|_\infty \leqq c\frac{\log n}{n}\omega_\varphi\Big(f^{(r)};\frac{1}{n}\Big), \quad \forall f \in C^r,\ 0 \leqq r \leqq m-2,$$

and

$$\|w_m(H_{mn}(w,f) - f)\|_\infty \leqq c\frac{\log n}{n^{m-1}}E_n(f^{(m-1)}), \quad \forall f \in C^{m-1}.$$

The proof of this theorem is similar to that of Theorem 2.1. One only needs to notice that the extra weight $w_m(x)$ will cancel out $w_m^{-1}(x)$ in (3.4)–(3.6). From this, it seems that the natural measurement for the uniform convergence of this type operator is the norm with suitable weight. See also [3].

The norm estimate of Hermite–Fejér interpolation for $m =$ even is given by $\omega_\varphi\left(f;\frac{\log n}{n}\right)$ ([14]), however, it seems unlikely that our estimate can be improved to $\frac{1}{n^r}\omega_\varphi\left(f^{(r)};\frac{\log n}{n}\right)$ for either $m =$ even or $m =$ odd. The reason is that the estimates are sharp in Lemmas 3.1 and 3.2 as shown by [11].

REFERENCES

[1] DITZIAN, Z. and TOTIK, V., *Moduli of smoothness*, Springer Series in Computational Mathematics, Vol. 9, Springer-Verlag, New York–Berlin, 1987. *MR* **89h**: 41002

[2] ESSER, H. and SCHERER, K., Eine Bemerkung zur Konvergenz Hermitescher Interpolationsprozesse, *Numer. Math.* **21** (1973), 220–222. *MR* **53** #6979

[3] HÁY, B. and VÉRTESI, P., Interpolation in spaces of weighted maximum norm, *Studia Sci. Math. Hungar.* **14** (1979), 1–9. *MR* **84d:** 41003

[4] MÁTÉ, A., NEVAI, P. and XU, Y., Mean convergence of Hermite interpolation (to appear).
[5] NEVAI, P., Mean convergence of Lagrange interpolation. III, *Trans. Amer. Math. Soc.* **282** (1984), 669–698. *MR* **85c**: 41009
[6] NEVAI, P., Bernstein's inequality in L^p for $0 < p < 1$, *J. Approx. Theory* **27** (1979), 239–243. *MR* **80m**: 41009
[7] NEVAI, P., Orthogonal polynomials, *Mem. Amer. Math. Soc.* **18** (1979), no. 213. *MR* **80k**: 42025
[8] NEVAI, P. and VÉRTESI, P., Mean convergence of Hermite–Fejér interpolation, *J. Math. Anal. Appl.* **105** (1985), 26–58. *MR* **86h**: 41004
[9] SZABADOS, J. and VARMA, A. K., On higher order Hermite–Fejér interpolation in weighted L^p-metric (to appear).
[10] VÉRTESI, P., Hermite–Fejér interpolations of higher order, I, *Acta Math. Hungar.* **54** (1989), 135–152. *MR* **90k**: 41008
[11] VÉRTESI, P., Hermite–Fejér interpolations of higher order, II, *Acta Math. Hungar.* **56** (1990), 369–380.
[12] VÉRTESI, P. and XU, Y., Order of mean convergence of Hermite–Fejér interpolation, *Studia Sci. Math. Hungar.* **24** (1989), 391–401.
[13] VÉRTESI, P. and XU, Y., Weighted L^p convergence of Hermite interpolation of higher order, *Acta Math. Hungar.* **59** (1992), 423–438.
[14] XU, Y., Rate of convergence of Hermite–Fejér interpolation of higher order, *J. Approx. Theory* (to appear).
[15] XU, Y., Note on simultaneous approximation by polynomials (to appear).
[16] XU, Y., On the norm of the Hermite interpolation operator, *Approximation theory* VI (College Station, TX, 1989), Vol. II, C. K. Chui et al., eds., Academic Press, Boston MA, 1989, 683–686. (See *MR* **91j**: 41002.)

(Received February 8, 1990)

MTA MATEMATIKAI KUTATÓINTÉZETE
P.O. BOX 127
H–1364 BUDAPEST
HUNGARY

DEPARTMENT OF MATHEMATICS
THE UNIVERSITY OF TEXAS AT AUSTIN
AUSTIN, TX 78712
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 215–227

SEMIGROUPS (RINGS) HAVING A PRIMITIVE REGULAR (COMPLETELY SEMISIMPLE) IDEAL

O. STEINFELD

A non-zero quasi-ideal Q of a semigroup with 0 (a ring) A is called *canonical* if Q is an intersection of a 0-minimal (minimal) right ideal R and a 0-minimal (minimal) left ideal L of A. This notion is due to A. H. Clifford [2]. In the papers [12], [11] we proved that the product of two canonical quasi-ideals of A is either 0 or a canonical quasi-ideal of A, that is, the set V of all canonical quasi-ideals of A and the zero element 0 of A form a multiplicative semigroup with 0 with respect to the multiplication of the quasi-ideals of A.

The first purpose of this paper is to answer the following question of R. P. Sullivan: "In which semigroups and rings is this multiplicative semigroup $V \cup 0$ regular?" In Theorem 1.4 we give several solutions of this problem. We mention here only one of them: $V \cup 0$ is a regular semigroup with 0 iff every canonical quasi-ideal $Q = R \cap L$ of A is such that R and L are (globally) idempotent. These canonical quasi-ideals of A will be called *distinguished.*

Our second aim is to give a simple joint characterization of the primitive regular (in particular, completely 0-simple) semigroups with 0 and of the completely semisimple (in particular, completely simple) rings by means of their distinguished canonical quasi-ideals (see Theorem 4.1 and Corollary 4.3).

§1. Several solutions of a problem of R. P. Sullivan

An additive subgroup (a non-empty subset) Q of a ring (a semigroup) A is called a *quasi-ideal* of A if $QA \cap AQ \subseteq Q$. A non-zero quasi-ideal Q of a semigroup with 0 (a ring) A is called *canonical* if Q is the intersection of a 0-minimal (minimal) right ideal R and a 0-minimal (minimal) left ideal L of A, that is, $0 \neq Q = R \cap L$. This important notion was introduced by A. H.

1991 *Mathematics Subject Classifications.* Primary 20M17; Secondary 16D30.

Key words and phrases. Quasi-ideal of a semigroup (ring), regular semigroup (ring), completely zero-simple semigroup, completely simple ring.

*Research supported by the Hungarian National Foundation for Scientific Research Grant No. 1813.

Akadémiai Kiadó, Budapest

Clifford [2] for semigroups with 0, and he has proved in [2] many interesting results concerning it.

REMARK 1.1. It is not difficult to prove the following proposition (see e.g. Lemma 6.27 in Clifford–Preston [4]): if R is an *idempotent* 0-minimal (minimal) right ideal of a semigroup with 0 (a ring) A and if L is a 0-minimal (minimal) left ideal of A such that $L^2 = 0$, then $R \cap L = 0$. (Evidently, the left-right dual of this result also holds.) On the other hand, Example 3.1 in Steinfeld–Thang [12] and Example 3.1 in Steinfeld [11] show that there exist semigroups with 0 and rings which have canonical quasi-ideals $Q = R \cap L$ such that $R^2 = L^2 = 0$. These facts mean that a canonical quasi-ideal $Q = R \cap L$ of A has the property either $R^2 = R$, $L^2 = L$ or $R^2 = L^2 = 0$.

We shall say that a *canonical quasi-ideal* $Q = R \cap L$ of A is *distinguished* if $R^2 = R$ and $L^2 = L$ hold.

REMARK 1.2. It is easy to show that every canonical quasi-ideal $Q = R \cap L$ of a semigroup with 0 (a ring) A is a 0-*minimal* (*minimal*) quasi-ideal of A (see e.g. Theorem 6.1 in our book [10]). But Example 7.1(a) in [10] shows that *not every* 0-minimal (minimal) quasi-ideal of A is canonical. From Theorems 6.3 and 6.5 of [10] it follows that a 0-minimal (minimal) quasi-ideal Q of A is either idempotent or $Q^2 = 0$. Hence a canonical quasi-ideal $Q = R \cap L$ of A is also either idempotent or a zero semigroup (zero ring); if $R^2 = L^2 = 0$, then evidently $Q^2 = 0$. It is evident that every canonical quasi-ideal $Q = R \cap L$ of a full matrix ring D_n of degree $n(\geq 2)$ over a division ring D is distinguished, but the case $Q^2 = 0$ is also possible.

In the papers Steinfeld–Thang [12] and Steinfeld [11] we have proved:

PROPOSITION 1.1 (see Corollary 2.7 in [12] and Corollary 2.6 in [11]). *If the semigroup with* 0 *(the ring)* A *contains at least one canonical quasi-ideal, then the union (the sum)* U *of all canonical quasi-ideals of* A *is a two-sided ideal of* A.

(Examples 3.6 and 3.2 in Steinfeld [10] show that the union (the sum) of two quasi-ideals of a semigroup (a ring) A is not always a quasi-ideal of A.)

PROPOSITION 1.2 (see Corollary 2.10 in [12] and Corollary 2.9 in [11]). *Assume that the semigroup with* 0 *(the ring)* A *has at least one canonical quasi-ideal. Let* V *denote the set of all canonical quasi-ideals of* A, *then* $V \cup 0$ *is a multiplicative semigroup with* 0 *with respect to the multiplication of the quasi-ideals of* A.

(Examples 3.7 and 3.8 in [10] show that the product of two quasi-ideals of a *semigroup* S is not always a quasi-ideal of S. Concerning rings, see Problems 3.1a and 3.1b in [10].)

PROBLEM of R. P. Sullivan. Characterize the semigroups with 0 (the rings) A in which the multiplicative semigroup $V \cup 0$ is regular.

In order to solve this problem we have to define some notions and we need some preliminaries.

A non-zero *idempotent element* e of a semigroup with 0 (a ring) A is called *primitive* if, for any non-zero idempotent f of A, the relation $ef = fe = f$ implies that $e = f$.

A *regular semigroup* S with 0 is said to be *primitive regular* if every non-zero idempotent of S is primitive. We shall say that an ideal M of a semigroup S with 0 is a *primitive regular ideal* of S if M is a primitive regular semigroup.

A *semigroup* S *with* 0 is called 0-*simple* if it has only two two-sided ideals $\{0\}$ and S, furthermore $S^2 \neq 0$. According to Theorem 2.48 of Clifford–Preston [3], by a *completely* 0-*simple semigroup* we shall mean a 0-simple semigroup S containing at least one 0-minimal left ideal and at least one 0-minimal right ideal. We shall say that an ideal M of a semigroup S with 0 is a *completely* 0-*simple ideal* of S if M is a completely 0-simple semigroup.

A *ring* A is called *simple* if it has only two two-sided ideals $\{0\}$ and A, furthermore $A^2 \neq 0$. In view of Corollary 5.4 B of Artin–Nesbitt–Thrall [1] (see also Proposition 2.5), by a *completely simple ring* we shall mean a simple ring having at least one minimal left ideal *or* at least one minimal right ideal. We shall say that an ideal M of a ring A is a *completely simple ideal* of A if M is a completely simple ring. Let a ring A be the discrete direct sum of its completely simple ideals, then A is said to be a *completely semisimple ring*. An ideal M of a ring A is called a *completely semisimple ideal* of A if M is a completely semisimple ring.

Let S be a semigroup with 0. The union of $\{0\}$ and of all the 0-minimal right ideals of S is called the *right socle* $\sum_r$ of S. The *left socle* $\sum_l$ of S is the union of $\{0\}$ and of all the 0-minimal left ideals of S. Theorem 6.29 of Clifford–Preston [4] gives the structure of the two-sided ideal $\sum_r \cup \sum_l$ of S; there is a strong analogy with Dieudonné's theory of the right and left socles of a ring. (See Dieudonné [5].) We shall say that S is the 0-*direct union* of the subsemigroups $\{S_i : i \in I\}$ if S is their 0-disjoint union and if $S_i S_j = S_j S_i = 0$ for $i \neq j$ $(i, j \in I)$.

THEOREM 1.3a (see Theorem 6.29 in Clifford–Preston [4]). *Let S denote a semigroup with* 0, *and let $C(C')$ be the union of* $\{0\}$ *and all those non-nilpotent* 0-*minimal right (left) ideals of S which have a non-zero intersection with some* 0-*minimal left (right) ideal of S. Then $C = C'$, and C is a two-sided ideal of S which is the* 0-*direct union of* $\{0\}$ *and all the completely* 0-*simple ideals of S.*

Let $\{A_i : i \in I\}$ denote a *non-empty* family of subrings of a ring A. Then the *subring* of A generated by $\bigcup_{i \in I} A_i$ is called the *sum of the subrings* $\{A_i : i \in I\}$. A sum of an *empty set* of subrings of A is defined to be equal to zero. The sum of all minimal right (left) ideals of a ring A is called the *right* (*left*)

socle of A. The following theorem is analogous with Theorem 1.3a. I think that it may be known, but I could not find it in the literature. Its proof is given in the next section.

THEOREM 1.3b. *Let A be a ring, and let $C(C')$ denote the sum of all those non-nilpotent minimal right (left) ideals of A which have a non-zero intersection with some minimal left (right) ideal of A. Then $C = C'$, and C is a two-sided ideal of A which is the discrete direct sum of all the completely simple ideals of A.*

REMARK 1.3. Let A be a semigroup with 0 (a ring). In the case when the union (the sum) $C = C'$ defined in Theorem 1.3a (Theorem 1.3b) is *not zero*, then we shall say that $C = C'$ is the *amicable part of the right and left socles of A.*

Now we are able to give the main result of this section.

THEOREM 1.4. *Assume that the semigroup with* 0 *(the ring) A has at least one canonical quasi-ideal. Then the following conditions are equivalent:*

(1) *the set V of all the canonical quasi-ideals of A together with* 0 *forms a regular multiplicative semigroup with* 0*;*

(2) *every canonical quasi-ideal of A is distinguished;*

(3) *the union (the sum) U of all canonical quasi-ideals of A equals to the amicable part $C = C'$ of the right and left socles of A;*

(4) *the union (the sum) U of all canonical quasi-ideals of A is a primitive regular ideal (a completely semisimple ideal) of A;*

(5) *A has a primitive regular ideal (a completely semisimple ideal) B which contains all the canonical quasi-ideals of A.*

We shall prove this theorem in Section 3.

At the end of this section we are going to prove a useful general relation.

PROPOSITION 1.5. *Let A be a semigroup with* 0 *(a ring) having at least one canonical quasi-ideal. Then the union (the sum) $C = C'$ defined in Theorem 1.3a (Theorem 1.3b) is always contained in the union (the sum) U defined in Proposition 1.1.*

PROOF. In the case when $C = C'$ is *zero*, our assertion is trivial. Now let $C = C'$ be *not zero*. Consider an arbitrary *non-nilpotent* 0-minimal (minimal) right ideal R of A contained in C. Then there exists a 0-minimal (minimal) left ideal L of A such that $R \cap L \neq 0$, whence $R \cap L \subseteq U$. Consider the product $(R \cap L)A$. Since R is a non-nilpotent 0-minimal (minimal) right ideal of A, it is easy to show that $(R \cap L)A$ is a *non-zero* right ideal of A contained in R, whence $(R \cap L)A = R$. On the other hand, by Proposition 1.1, U is an ideal of A, so we have $R = (R \cap L)A \subseteq UA \subseteq U$. This inclusion and Theorem 1.3a (Theorem 1.3b) imply that the union (the sum) $C = C'$ is always contained in the union (the sum) U, in fact.

REMARK 1.4. In Remark 1.1 we have mentioned that there exist rings and semigroups with 0 which have *canonical* quasi-ideals $Q = R \cap L$ such that $R^2 = 0$ and $L^2 = 0$. Since also these canonical quasi-ideals are contained in U, but they are *not* contained in $C = C'$, the inclusion $U \subseteq C = C'$ is *not always* true.

§2. Preliminaries

In order to prove Theorems 1.3b and 1.4 we need some known or partially known results.

LEMMA 2.1 (see Theorems 1.3 and 1.7 in Gluskin–Steinfeld [6] and the proofs of Theorems 3.2 and 3.3 in the papers [12] and [11], respectively). *Let A be a semigroup with 0 (a ring). If Q is an idempotent canonical quasi-ideal of A, then AQA is a completely 0-simple (completely simple) ideal of A. If R and L are 0-minimal (minimal) right and left ideals of A, respectively, such that $RL \neq 0$, then $ARLA$ is also a completely 0-simple (completely simple) ideal of A.*

REMARK 2.1. It is easy to show that the product $RL(\neq 0)$ is also an idempotent canonical quasi-ideal of A.

LEMMA 2.2a (see Theorem 6.39 in Clifford–Preston [4] and Theorem 10.1 in Steinfeld [10]). *The following conditions on a semigroup S with 0 are equivalent:*

(1) *S is primitive regular;*

(2) *S is the 0-direct union of its completely 0-simple ideals;*

(3) *S is regular and the union of its 0-minimal quasi-ideals.*

LEMMA 2.2b (cf. Theorem 78.2 in Kertész [7], Theorem 8.1 in Steinfeld [10] and Szász [13]). *The following conditions on a ring A are equivalent:*

(a) *A is completely semisimple, i.e. A is the discrete direct sum of its completely simple ideals;*

(b) *A is regular and the sum of its minimal right ideals;*

(c) *A is regular and the sum of its minimal quasi-ideals.*

REMARK 2.2. It is known that conditions (a) and (b) characterize the semisimple rings satisfying the minimum condition for principal right ideals (see Szász [13], Kertész [7]).

LEMMA 2.3 (cf. Lemma 6.38 in Clifford–Preston [4] and Lemmas 10.3a and 10.3b in our book [10]). *Let A be a ring (semigroup with 0) and let e be a non-zero idempotent element of A, then the following conditions are equivalent:*

(1) *e is a primitive idempotent element, and every element of the left ideal Ae is regular;*

(2) *Ae is a minimal (0-minimal) left ideal of A.*

PROPOSITION 2.4 (see e.g. van der Waerden [14]). *Any minimal right ideal R of a ring A is either a zero ring or it has a non-zero idempotent element e such that $R = eA$.*

PROPOSITION 2.5 (cf. Corollary 5.4B in Artin–Nesbitt–Thrall [1]). *Let e be a non-zero idempotent element of a simple ring A. Then Ae is a minimal left ideal of A if and only if eA is a minimal right ideal of A.*

Now we shall prove the following important

THEOREM 2.6 (cf. Rich [9] and Gluskin–Steinfeld [6]). *If M is a completely simple (completely 0-simple) ideal of a ring (semigroup with 0) A, then for every minimal (0-minimal) right ideal R of M there exists a minimal (0-minimal) left ideal L of M such that $M = LR$ and $RL \neq 0$, and for every minimal (0-minimal) left ideal L' of M there exists a minimal (0-minimal) right ideal R' of M such that $M = L'R'$ and $R'L' \neq 0$. Furthermore any minimal (0-minimal) right and left ideals of M are minimal (0-minimal) right and left ideals of A, respectively.*

PROOF. *First let A be a ring* and let M be a completely simple ideal of A. In view of Lemma 2.2b, M is a *regular* ring, furthermore from Proposition 2.4 it follows that every minimal right ideal R of M has the form

$$R = eM \qquad (0 \neq e^2 = e \in M). \tag{2.1}$$

By Proposition 2.5, the left ideal $L = Me$ of the *simple* ring M is minimal. The simplicity of M implies that $LR = Me^2M = M$, furthermore $RL = eM \cdot Me \neq 0$. From the dual of Proposition 2.4, it follows that every minimal left ideal L' of M has the form

$$L' = Mf \qquad (0 \neq f^2 = f \in M). \tag{2.2}$$

Again by Proposition 2.5, we get that the right ideal $R' = fM$ of M is minimal such that $L'R' = Mf^2M = M$ and $R'L' = fM \cdot Mf \neq 0$.

From the relations (2.1) and (2.2), it follows immediately that any minimal right (left) ideal of M is a minimal right (left) ideal of A.

Now let A be a semigroup with 0 and let M denote a completely 0-simple ideal of A. In view of Lemma 2.2a, M is a regular semigroup with 0. Since any 0-minimal right ideal R of M is generated by any non-zero element r of R and there exists an element m in M such that $r = rmr$, we get that for the idempotent element $rm = e$ of R it holds $R = eM$ $(0 \neq e^2 = e \in M)$. Again by Lemma 2.2a, the idempotent element e of M is *primitive* and every element of the left ideal $Me = L$ of M is *regular*. From Lemma 2.3 it follows that $L = Me$ is a 0-minimal left ideal of M. Since M is a 0-simple semigroup, we have that $LM = Me^2M = M$ and $RL = eMMe \neq 0$. If we consider an arbitrary 0-minimal left ideal L' of M, then one can conclude by the dual

of the foregoing argument that there exists an idempotent element $f(\neq 0)$ in M such that $L' = Mf$. By Lemma 2.2a and by the dual of Lemma 2.3 one gets that $fM = R'$ is a 0-minimal right ideal of M. The proof can be continued as above.

Finally, one can show immediately that any 0-minimal right and left ideals of M are 0-minimal right and left ideals of A, respectively.

Now we are able to begin the

PROOF of Theorem 1.3b. Let R be a *non-nilpotent* minimal right ideal of a ring A such that there exists some minimal left ideal L of A such that $R \cap L \neq 0$. Since R must be *idempotent* and the intersection $R \cap L$ is *not zero*, from proposition given in Remark 1.1 it follows that the left ideal L of A is also *idempotent*, that is, $R^2 = R$, $L^2 = L$ and $R \cap L \neq 0$ hold. Proposition 2.4 and its dual imply that there exist non-zero idempotent elements e and f in A such that $R = eA$ and $L = Af$, and evidently

$$0 \neq R \cap L = eA \cap Af = eAf.$$

Since Af is a minimal left ideal of A and $0 \neq eAf \subseteq A^2 f \subseteq Af$ holds, we have that $A^2 f = Af$, whence $eA \cdot Af = eAf \neq 0$. This relation and Lemma 2.1 imply that $ARLA = AeAAfA = K$ is a *completely simple ideal* of A. Since $0 \neq RLA$ is a right ideal of A contained in the minimal right ideal R of A, we have that $RLA = R$, whence $R = R^2 = RRLA \subseteq ARLA = K$. From the definition of C given in Theorem 1.3b and from $R \subseteq K$ it follows that C is contained in the discrete direct sum C^* of all the completely simple ideals of A.

Conversely, let M be a completely simple ideal of A and let R be a (non-nilpotent) minimal right ideal of M. By Theorem 2.6, there exists a minimal left ideal L of M such that $0 \neq RL \subseteq R \cap L$. Again by Theorem 2.6, R and L are minimal right and left ideals of A, respectively. These facts imply that $R \subseteq C$. On the other hand, by Lemma 2.2b, the completely simple ring M is the sum of its minimal right ideals. This property and $R \subseteq C$ imply that $M \subseteq C$. Thus the discrete direct sum C^* of all the completely simple ideals of A is contained in C. We conclude that $C^* = C$.

By the left-right duality of the foregoing reasoning one gets that $C^* = C'$, which completes our proof.

§3. Proof of Theorem 1.4

First let A be a semigroup with 0.

(1) $\Longrightarrow$ (2). Let Q_1 be an arbitrary canonical quasi-ideal of A. By condition (1), there exists a canonical quasi-ideal Q_2 of A such that $Q_1 Q_2 Q_1 = Q_1$. Since $Q_1 Q_2 = Q_3$ is an *idempotent* canonical quasi-ideal of A, from Lemma 2.1 it follows that AQ_3A is a completely 0-simple ideal of A. Evidently it holds

$$Q_1 = (Q_1 Q_2)Q_1 = Q_3 Q_1 = Q_3^2 Q_1 \subseteq AQ_3A.$$

On the other hand, the canonical quasi-ideal Q_1 of A has the form $Q_1 = R_1 \cap L_1$, where R_1 and L_1 are 0-minimal right and 0-minimal left ideals of A, respectively. Consider an arbitrary non-zero element x of Q_1, then x belongs to R_1. In view of Lemma 2.2a, the completely 0-simple ideal AQ_3A of A is a *regular* semigroup, thus we have that

$$R_1 = xA \subseteq (AQ_3A)A \subseteq AQ_3A,$$

whence $R_1^2 = R_1$. Dually we get that $L_1^2 = L_1$. These mean that $Q_1 = R_1 \cap L_1$ is a *distinguished* canonical quasi-ideal of A, in fact.

(2) $\Longrightarrow$ (3). By Proposition 1.5, the union $C = C'$ is always contained in the union U. In order to show the inclusion $U \subseteq C = C'$, consider an arbitrary canonical quasi-ideal $Q = R \cap L$ of A. By condition (2), the 0-minimal right ideal R and the 0-minimal left ideal L of A are *not nilpotent*. These facts and $Q = R \cap L \neq 0$ imply that $R \subseteq C$ and $L \subseteq C'$. From these inclusions and from Theorem 1.3a it follows that $Q = R \cap L \subseteq C = C'$. Since U is the union of all the canonical quasi-ideals of A, we conclude that $U \subseteq C = C'$ indeed.

(3) $\Longrightarrow$ (4). Assume that the equality $U = C = C'$ holds, then Theorem 1.3a and Lemma 2.2a imply that U is a primitive regular ideal of A, in fact.

The implication (4) $\Longrightarrow$ (5) is trivial.

(5) $\Longrightarrow$ (1). Consider an arbitrary canonical quasi-ideal $Q = R \cap L$ of A, where R and L are 0-minimal right and 0-minimal left ideals of A, respectively. By condition (5), A has a *primitive regular ideal* B which contains Q. Let q be an arbitrary *non-zero* element of $Q = R \cap L \subseteq B$. Since $q(\in B)$ is a non-zero *regular* element, and it belongs to the 0-minimal right ideal R and 0-minimal left ideal L of A, we have that

$$R = qA \quad \text{and} \quad L = Aq \qquad (0 \neq q \in Q \subseteq B).$$

The regular element $q(\in B)$ has the form $q = qxq$ for some x in B, then $qx = e$ and $xq = f$ are idempotent elements in B such that

$$R = eA \quad \text{and} \quad L = Af \qquad (0 \neq e^2 = e,\ 0 \neq f^2 = f),$$

whence $Q = R \cap L = eA \cap Af = eAf$. Since the elements e and f belong to the ideal B, not only $R = eA$ and $L = Af$, but also Ae and fA are contained in B. By condition (5), the idempotent elements e and f of B are *primitive*, furthermore every element of the left ideal Ae and every element of the right ideal fA are *regular*. These facts, Lemma 2.3 and its dual imply that Ae and fA are 0-minimal left and 0-minimal right ideals of A, respectively.

The intersection $fA \cap Ae = fAe$ is either 0 or a canonical quasi-ideal of A. In order to show that fAe is *not zero*, consider a non-zero element eaf of the canonical quasi-ideal $Q = eAf$ of A. Since Af is a 0-minimal left ideal of A and $0 \neq AeAf \subseteq Af$ holds, we get $Aeaf = Af$, whence $fAeaf = fAf$. This

relation implies that there exists at least one *non-zero element* fbe in fAe such that $fbeeaf = f$. We can conclude that fAe is a *canonical quasi-ideal* of A such that

$$fAeeAf = fAf.$$

Premultiplying this equation by eAf we have that

$$eAffAeeAf = eAffAf \subseteq eAf.$$

Evidently, $eAffAf$ is a *non-zero quasi-ideal* of A contained in the 0-*minimal* quasi-ideal eAf of A, so we have that

$$eAffAeeAf = eAf.$$

This relation means that $V \cup 0$ is a *regular multiplicative semigroup with* 0, in fact.

Now let A be a ring. The proof of the implications (1) $\Longrightarrow$ (2), (2) = $\Longrightarrow$ (3), (3) $\Longrightarrow$ (4), (4) $\Longrightarrow$ (5) and (5) $\Longrightarrow$ (1) runs analogously as in the case of semigroups, if we apply Theorem 1.3b and Lemma 2.2b instead of Theorem 1.3a and Lemma 2.2a.

§4. On primitive regular semigroups and on completely semisimple rings

By means of Theorem 1.4 it would be possible to characterize the primitive regular semigroups and the completely semisimple rings, but these characterizations would be too complicated.

In view of Lemmas 2.2a and 2.2b, a semigroup with 0 (a ring) A is primitive regular (completely semisimple) if and only if A is *regular* and the union (the sum) of its 0-minimal (minimal) quasi-ideals. In Theorem 4.1 we can get rid of the requirement of regularity by imposing two simple additional properties of these 0-minimal (minimal) quasi-ideals of A. (Cf. condition (D) in Theorem 10.1 and condition (C) in Theorem 8.1 of Steinfeld [10]; see also Section 1 in Márki [8].)

THEOREM 4.1. *Let A be a semigroup with* 0 *(a ring). A is primitive regular (completely semisimple) if and only if A is the union (the sum) of its distinguished canonical quasi-ideals.*

For the proof we need the following

PROPOSITION 4.2 (see Corollaries 7.3a and 7.3b in Steinfeld [10]). *Let A be a regular semigroup with* 0 *(a regular ring). Every* 0-*minimal (minimal) quasi-ideal Q of A is distinguished canonical, more precisely, every Q has the form* $Q = eA \cap Af$ $(0 \neq e^2 = e \in A,\ 0 \neq f^2 = f \in A)$, *where eA and Af are idempotent* 0-*minimol (minimal) right and left ideals of A, respectively.*

PROOF of Theorem 4.1. *First let A be a semigroup with* 0. Assume that A is primitive regular. By Lemma 2.2a, A is the union of its 0-minimal

quasi-ideals, which are, in view of Proposition 4.2, distinguished canonical quasi-ideals of A.

Conversely, assume that A is the union of its distinguished canonical quasi-ideals $Q_{\gamma\delta} = R_\gamma \cap L_\delta$ $(\gamma \in \Gamma, \delta \in \Delta)$, where R_γ and L_δ are idempotent 0-minimal right and left ideals of A, respectively. Hence

$$A = \bigcup_{\gamma\in\Gamma} R_\gamma = \bigcup_{\delta\in\Delta} L_\delta \qquad (R_\gamma^2 = R_\gamma,\ L_\delta^2 = L_\delta). \tag{4.1}$$

Consider an arbitrary non-zero element a of A. In view of (4.1), the element a belongs to an idempotent 0-minimal left ideal $L_\delta = L$ $(\delta \in \Delta)$ of A, whence $Aa \subseteq L$. The 0-minimality of L implies that either $Aa = 0$ or $Aa = L$. We shall show that Aa is *not zero*. From the assumption $Aa = 0$ $(0 \neq a \in L)$ it follows namely that the set M of all elements a of L such that $Aa = 0$ is a *non-zero* left ideal of A contained in L. By the 0-minimality of L, we have $M = L$, that is, $AM = AL = 0$, whence $L^2 = 0$, what is a contradiction. So we get that $Aa = L$. This relation implies the existence of an element e in A such that $ea = a$. Again by (4.1), the element e belongs to an idempotent 0-minimal right ideal $R_\gamma = R$ $(\gamma \in \Gamma)$ of A, whence $a = ea \in R$. Since R is a 0-minimal right ideal of A, either $aA = 0$ or $aA = R$ holds. Dually as above we get that $aA = R$ must hold. So there exists an element x in A such that $ax = e$. This equation and $ea = a$ imply that $a = axa$, that is, A is a *regular* semigroup.

On the other hand, let f denote an arbitrary non-zero idempotent element of A. Then f belongs to a 0-minimal left ideal $L_{\delta'} = L'(\delta' \in \Delta)$ of A, whence $L' = Af$. By Lemma 2.3, the idempotent element f of A is *primitive*. Q. e. d.

Now let A be a ring. Assume that it is completely semisimple, that is, A is the discrete direct sum of its completely simple ideals. By Lemma 2.2b, A is a *regular ring* and the sum of its minimal quasi-ideals. In view of Proposition 4.2, every minimal quasi-ideal of A is distinguished canonical.

Conversely, assume that A is the sum of its distinguished canonical quasi-ideals $Q_{\gamma\delta} = R_\gamma \cap L_\delta$ $(\gamma \in \Gamma, \delta \in \Delta)$, where the R_γ and L_δ are *idempotent* minimal right and left ideals of A, respectively. Hence

$$A = \sum_{\gamma\in\Gamma} R_\gamma = \sum_{\delta\in\Delta} L_\delta \qquad (R_\gamma^2 = R_\gamma,\ L_\delta^2 = L_\delta). \tag{4.2}$$

Consider a fixed idempotent minimal right ideal $R_{\gamma^*} = R$ $(\gamma^* \in \Gamma)$ of A. We have that

$$RA = R\Big(\sum_{\gamma\in\Gamma} R_\gamma\Big) = \sum_{\gamma\in\Gamma} (RR_\gamma) \neq 0. \tag{4.3}$$

In view of (4.2) and (4.3) one gets that

$$0 \neq RA = R\Big(\sum_{\delta\in\Delta} L_\delta\Big) = \sum_{\delta\in\Delta} RL_\delta.$$

From this relation it follows that for any fixed idempotent minimal right ideal $R_{\gamma^*} = R$ $(\gamma^* \in \Gamma)$ of A there exists at least one minimal left ideal $L_{\delta^*} = L$ $(\delta^* \in \Delta)$ of A such that $RL \neq 0$. By Lemma 2.1, the product $ARLA = B$ is a *completely simple ideal* of A. Since $RL \neq 0$ and $L^2 = L$ hold, we have that

$$0 \neq RL = RL^2 \subseteq RLA \subseteq R,$$

whence $RLA = R$. This relation and $R^2 = R$ imply that

$$R_{\gamma^*} = R = R^2 = RRLA \subseteq ARLA = B \qquad (\gamma^* \in \Gamma).$$

This relation means that any idempotent minimal right ideal $R_{\gamma^*} = R$ $(\gamma^* \in \Gamma)$ of A is contained in a suitable completely simple ideal $B = ARLA$ of A. From this fact and from (4.2) it follows that the ring $A = \sum_{\gamma \in \Gamma} R_\gamma$ is just the discrete direct sum of its completely simple ideals, that is, A is a completely semisimple ring, in fact.

Theorem 4.1 has the following

COROLLARY 4.3. *Let A be a semigroup with 0 (a ring). A is completely 0-simple (completely simple) if and only if it is the union (the sum) of its canonical quasi-ideals $Q_{\gamma\delta} = R_\gamma \cap L_\delta$ $(\gamma \in \Gamma, \delta \in \Delta)$ such that*

$$R_\gamma R_{\gamma'} = R_\gamma \text{ for all } \gamma, \gamma' \in \Gamma \text{ and } L_\delta L_{\delta'} = L_{\delta'} \text{ for all } \delta, \delta' \in \Delta. \tag{4.4}$$

PROOF. *First let A be a semigroup with 0.* Assume that A is completely 0-simple. By Lemma 2.2a, A is primitive regular. From Theorem 4.1 it follows that A is the union of its canonical quasi-ideals $Q_{\gamma\delta} = R_\gamma \cap L_\delta$ $(\gamma \in \Gamma, \delta \in \Delta)$. For any two 0-minimal right ideals R_γ, $R_{\gamma'}$ $(\gamma, \gamma' \in \Gamma)$ of A it holds evidently

$$\text{either } R_\gamma R_{\gamma'} = 0 \text{ or } R_\gamma R_{\gamma'} = R_\gamma.$$

Since A is 0-simple, $\{0\}$ is a prime ideal of A, therefore the case $R_\gamma R_{\gamma'} = 0$ is not possible.

One gets dually that $L_\delta L_{\delta'} = L_{\delta'}$ must hold for all $\delta, \delta' \in \Delta$.

Conversely, assume that A is the union of its canonical quasi-ideals $Q_{\gamma\delta} = R_\gamma \cap L_\delta$ $(\gamma \in \Gamma, \delta \in \Delta)$ satisfying condition (4.4). In view of Theorem 4.1, A is a primitive regular semigroup, so we have to show that A is 0-simple. Let B be a non-zero (two-sided) ideal of A. Since A is regular, B^2 is not zero. Hence

$$0 \neq B^2 \subseteq BA = B\Big(\bigcup_{\gamma \in \Gamma} \bigcup_{\delta \in \Delta} (R_\gamma \cap L_\delta)\Big) \subseteq \bigcup_{\delta \in \Delta} BL_\delta.$$

This relation implies that for some $\delta^* \in \Delta$ the product BL_{δ^*} is not zero. As BL_{δ^*} $(\delta^* \in \Delta)$ is a non-zero left ideal of A contained in the 0-minimal left ideal L_{δ^*} of A, we have that

$$L_{\delta^*} = BL_{\delta^*} \subseteq BA \subseteq B.$$

This relation and condition (4.4) imply that

$$A = \bigcup_{\gamma\in\Gamma}\bigcup_{\delta\in\Delta}(R_\gamma \cap L_\delta) \subseteq \bigcup_{\delta\in\Delta} L_\delta = \bigcup_{\delta\in\Delta} L_{\delta^*}L_\delta = L_{\delta^*}(\bigcup_{\delta\in\Delta} L_\delta) \subseteq BA \subseteq B,$$

that is, A is 0-simple, indeed.

Now let A be a ring. The proof runs analogously as in the case of semigroups, but we have to use Lemma 2.2b instead of Lemma 2.2a.

REMARK 4.1. Condition (4.4) implies that the minimal right ideals R_γ ($\gamma \in \Gamma$) and the minimal left ideals L_δ ($\delta \in \Delta$) of the *ring* A are (globally) *idempotent.* By Proposition 2.4 and its dual, every minimal right ideal R_γ ($\gamma \in \Gamma$) (minimal left ideal L_δ ($\delta \in \Delta$)) of A is generated by an *idempotent element.*

It is easy to prove the following proposition (see Proposition 6.12a in Steinfeld [10]): *Let e, f be non-zero idempotent elements of an arbitrary ring A such that eA, fA (Ae, Af) are minimal right (left) ideals of A. Then eA, fA (Ae, Af) are A-isomorphic right A-modules (left A-modules) iff the quasi-ideal eAf is not zero.*

These relations imply that the minimal right ideals R_γ ($\gamma \in \Gamma$) (minimal left ideals L_δ ($\delta \in \Delta$)) of A satisfying (4.4) are pairwise A-isomorphic right A-modules (left A-modules). Corollary 4.3 is a much simpler characterization of the completely simple rings than condition (C*) in Theorem 8.8 of Steinfeld [10].

REMARK 4.2. Proposition 2.4 and its dual are not true for semigroups, in general, but if a semigroup A with 0 is the union of its canonical quasi-ideals $Q_{\gamma\delta} = R_\gamma \cap L_\delta$ ($\gamma \in \Gamma$, $\delta \in \Delta$) satisfying condition (4.4), then Theorem 4.1 and Proposition 4.2 imply that every 0-minimal right ideal R_γ ($\gamma \in \Gamma$) (0-minimal left ideal L_δ ($\delta \in \Delta$)) of A is generated by an *idempotent element of A.* Since Proposition 6.12b in [10] is a semigroup theoretical analogue of Proposition 6.12a in [10] (see Remark 4.1), the mentioned relations and Proposition 6.12b in [10] imply that the 0-minimal right ideals R_γ ($\gamma \in \Gamma$) (0-minimal left ideals L_δ ($\delta \in \Delta$)) of A satisfying (4.4) are pairwise *right similar* (*left similar*) (the definitions are given on page 23 in [10]). Corollary 4.3 is a simpler characterization of the completely 0-simple semigroups than condition (D*) in Theorem 10.10 of [10].

REMARK 4.3. Finally we mention the following corollary of Theorem 4.1: Let S be a semigroup with 0. Then S is an inverse semigroup in which every non-zero idempotent is primitive iff S is the union of its distinguished canonical quasi-ideals $Q_{\gamma\delta} = R_\gamma \cap L_\delta$ ($\gamma \in \Gamma$, $\delta \in \Delta$) such that for every 0-minimal right ideal R_γ ($\gamma \in \Gamma$) there exists a unique 0-minimal left ideal $L_{\delta(\gamma)}$ ($\delta(\gamma) \in \Delta$) of S so that $R_\gamma \cap L_{\delta(\gamma)}$ is a group with 0. (Cf. condition (d) in Corollary 10.9 of Steinfeld [10].)

REFERENCES

[1] ARTIN, E., NESBITT, C. J. and THRALL, R. M., *Rings with minimum condition*, Univ. of Michigan Press, Ann Arbor, 1944. *MR* **6**–33

[2] CLIFFORD, A. H., Remarks on 0-minimal quasi-ideals in semigroups, *Semigroup Forum* **16** (1978), 183–196. *MR* **58** #6002

[3] CLIFFORD, A. H. and PRESTON, G. B., *The algebraic theory of semigroups* I, Mathematical Surveys, No. 7, American Mathematical Society, Providence, R.I., 1961. *MR* **24** #A2627

[4] CLIFFORD, A. H. and PRESTON, G. B., *The algebraic theory of semigroups* II, Mathematical Surveys, No. 7, American Mathematical Society, Providence, R.I., 1967. *MR* **36** # 1558

[5] DIEUDONNÉ, J., Sur le socle d'un anneau et les anneaux simples infinis, *Bull. Soc. Math. France* **70** (1942), 46–75. *MR* **6**–144

[6] GLUSKIN, L. M. and STEINFELD, O., Rings (semigroups) containing minimal (0-minimal) right and left ideals, *Publ. Math. Debrecen* **25** (1978), 275–280. *MR* **80c**: 20090

[7] KERTÉSZ, A., *Lectures on Artinian rings* (edited by R. Wiegandt), Disquisitiones Mathematicae Hungaricae, Vol. 14, Akadémiai Kiadó, Budapest, 1987. *MR* **88m**: 16016

[8] MÁRKI, L., A note on quasi-ideals and 0-matrix decompositions of semigroups with zero, Math. Inst. Hungar. Acad. Sci., Budapest, Sept. 1982 (preprint).

[9] RICH, R. P., Completely simple ideals of a semigroup, *Amer. J. Math.* **71** (1949), 883–885. *MR* **11**–327

[10] STEINFELD, O., *Quasi-ideals in rings and semigroups*, Disquisitiones Mathematicae Hungaricae, 10, Akadémiai Kiadó, Budapest, 1978. *MR* **80e**: 16001

[11] STEINFELD, O., On canonical quasi-ideals in rings, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **31** (1988), 171–178. *MR* **90f**: 16022

[12] STEINFELD, O. and THANG, T. T., Remarks on canonical quasi-ideals in semigroups, *Beiträge Algebra Geom.* **26** (1988), 127–135. *MR* **89f**: 20074

[13] SZÁSZ, F., Über Ringe mit Minimalbedingung für Hauptrechtsideale II, *Acta Math Acad. Sci. Hungar.* **12** (1961), 417–439. *MR* **26** #6207

[14] WAERDEN, B. L. VAN DER, *Algebra*, II, Springer-Verlag, Berlin, 1955. *MR* **17**–338

(Received June 15, 1990)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 229

BOOK REVIEW

Csörgő, M. and **Horváth, L.**, Weighted approximations in probability and statistics, John Wiley & Sons, Ltd., 1993. ISBN 0471936359.

The properties of the empirical distribution played an important role in mathematical statistics since the earliest time. The limit distribution of the distance between the empirical – and the real distribution functions was given by Kolmogorov in 1933. Already Kolmogorov himself realized that his statistic did not give too much information about the tails of the underlying distribution and proposed to investigate a weighted statistic instead. Rényi in 1953 was the first one who solved this problem using a special weight. Since then a number of results were published on the weighted empirical process using more and more general weights.

In 1949 Doob observed that the properties of the empirical process imitate those of a Brownian bridge. This observation initiated a new direction of research. Within this direction the best result is the so-called Hungarian construction.

The authors intend to give a complete overview of the properties of the weighted empirical and quantile processes via the Hungarian construction. They succeeded. In fact they can prove so-strong approximation results which can produce not only the known results of this theory but a lot of new theorems. One can say that having this book it is hard to find any new question on these weighted processes.

Beside the detailed study of the weighted processes the authors devote a chapter of their book to the renewal process. They prove that a renewal process can be also approximated by a Wiener process and they show how this approximation can be used to investigate the properties of a renewal process via having the corresponding properties of a Wiener process.

The only thing that bothers me is the numbering. For example it is very disturbing that the Theorems are numbered by an other system than the formulas.

P. Révész (Vienna)

Akadémiai Kiadó, Budapest

Typeset by TypoTEX Ltd., Budapest
PRINTED IN HUNGARY
Akadémiai Kiadó és Nyomda Vállalat, Budapest

RECENTLY ACCEPTED PAPERS

ASLAM, M., Matrix equations in radicals

KOVÁCS, K., On a generalization of an old theorem of Erdős

BOGNÁR, G., Eigenvalue problem for some nonlinear elliptic partial differential equation

WINKLER, R., Polynomial approximation on locally compact abelian groups, II

SHAO, Q.-M.,Random increments of a Wiener process and their applications

RISKIN, A., The crossing number of a cubic plane polyhedral map plus an edge

SLEZÁK, B., An inverse function theorem in topological groups

OHYA, M. and PETZ, D., Notes on quantum entropy

ROOS, C. and WIEGANDT, R., On the radical theory of graded rings which are inversive hemirings

LIANG, Z. and HE, Y., The Hun semigroup structure of point processes

KOMJÁTH, P., Partition of vector spaces

GRYTCZUK, A., On Fermat's equation in the set of integral 2×2 matrices

DEÁK, J., Spaces from pieces, II–IV

RÉVÉSZ, Sz. GY., On Beurling's prime number theorem

VECCHIA, B. D., MASTROIANNI, G. and VÉRTESI, P., Weighted L^p-approximation by Hermite interpolation of higher order plus end points

LIU, Z., On elliptic systems with discontinuous nonlinearities

CONTENTS

GONCHIGDORZH, R., Generalized p.p. rings and rings of π-regular quotients .. 1
KENT, R. E., Dialectical logic: the process calculus 17
БЕРМАН, Д. Л., К теории экстремальных полиномиальных операторов 63
KHAN, L. A., Integration of vector-valued continuous functions and the Riesz representation theorem 71
BIHARI, I., A generalization of the Riccati equation 79
SAKAI, R., and VÉRTESI, P., Hermite–Fejér interpolations of higher order. III 87
SOLTAN, V. P. and NGUEN, M. H., Lower bounds for the numbers of extremal and exposed diameters of a convex body 99
DEÁK, J., Extending a quasi-metric 105
FÉNYES, T., On an algebraic differential equation of Bernoulli type 113
WALENDZIAK, A., Join decompositions in lower continuous lattices 129
WINKLER, R., Polynomial approximation on locally compact abelian groups . 133
JOOS, K., Nonuniform convergence rates in the central limit theorem for martingales 143
BELL, H. E. and KLEIN, A. A., Two commutativity problems for rings 157
VÁSÁRHELYI, É., Covering of a triangle by homothetic triangles 161
KOMLÓS, J., REJTŐ, L. and TUSNÁDY, G., Learning with finite memory 171
SEBESTYÉN, Z., Restrictions of adjoint operators in Hilbert space 177
KY, N. X., On approximation by trigonometric polynomials in L_u^p-spaces 181
FÉNYES, T., On the Fourier transform of the modified Bessel function with respect to the order 187
FÉNYES, T., On the Fourier transform of the Bessel function with respect to the order 195
VÉRTESI, P. and XU, Y., Truncated Hermite interpolation polynomials 203
STEINFELD, O., Semigroups (rings) having a primitive regular (completely semisimple) ideal 213
BOOK REVIEW 227

HU ISSN 0081-6906

Studia Scientiarum Mathematicarum Hungarica

VOLUME 28
NUMBERS 3–4
1993

AKADÉMIAI KIADÓ, BUDAPEST

STUDIA SCIENTIARUM MATHEMATICARUM HUNGARICA

A QUARTERLY OF THE HUNGARIAN ACADEMY OF SCIENCES

Studia Scientiarum Mathematicarum Hungarica publishes original papers on mathematics mainly in English, but also in German, French and Russian. It is published in yearly volumes of four issues (mostly double numbers published semiannually) by

AKADÉMIAI KIADÓ
Publishing House of the Hungarian Academy of Sciences
H–1117 Budapest, Prielle Kornélia u. 19–35

Manuscripts and editorial correspondence should be addressed to

J. Merza
Managing Editor

P.O. Box 127
H–1364 Budapest

Tel.: (36)(1) 118–2875 Fax: (36)(1) 117–7166
e-mail: h3299mer @ ella.hu

Subscription information

Subscription price for Volume 28 (1993) in 4 issues: $ 88.00 including normal postage, airmail delivery plus $ 20.00

Orders should be addressed to

AKADÉMIAI KIADÓ
P.O.Box 245
H–1519 Budapest

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 231–236

ON AN APPLICATION OF A BINOMIAL SERIES EXPANSION OF DISCRETE OPERATORS

T. FÉNYES

Introduction

Let us consider the discrete Mikusiński operator field M_D based on the number-theoretical Dirichlet product of functions defined on the set of the positive integers. In the paper [1] we have discussed the algebraic Bernoulli-equation

$$D(x) + ax + bx^m = 0 \qquad x \in M_D \tag{1}$$

in M_D for integer values of m and briefly showed that the discussion can be easily extended to rational values of m. In (1) $a, b \in E$, where E denotes the ring of functions with the ordinary addition and with the Dirichlet product

$$fg = \left\{ \sum_{\nu|n} f(\nu) g\left(\frac{n}{\nu}\right) \right\} \qquad f, g \in E, \quad n = 1, 2, \dots,$$

moreover D denotes the well-known algebraic derivative (see [1]).

In this paper we shall deal with (1) if m is irrational. Irrational powers of the elements of M_D cannot be defined in general. However, we shall show that for every $h \in E$, $h(1) > 0$, and for every real number r, h^r can be defined by an operationally convergent binomial series. Let $\bar{E} \subset E$ be the subset of E, such that for every $h \in \bar{E}$, $h(1) > 0$.

So (1) can be defined for $x \in \bar{E}$ and for irrational values of m. By applying the results of paper [1] we prove simple existence criteria for the solutions of the Bernoulli equation and give the explicit form of the solutions belonging to $\bar{E}$.

The reader can find the elements of the discrete operational calculus in the paper [1]. For the applied notations used in this paper we refer also to [1].

1991 *Mathematics Subject Classifications.* Primary 44A40; Secondary 11A99, 13N99.
Key words and phrases. Operational calculus, number theory.

*Research partially supported by the Hungarian National Foundation for Scientific Research Grant No. 6032/6319.

Akadémiai Kiadó, Budapest

§ 1. On a binomial series expansion of discrete operators

We state the following

THEOREM 1. *Let $f \in E$, $f(1) = 0$, and let r be an arbitrary real number. Then*

$$e^{r \int \frac{D(f)}{1+f}} = \sum_{\nu=0}^{\infty} \binom{r}{\nu} f^{\nu} \tag{1.1}$$

in the sense of the pointwise convergence for every fixed n.

PROOF. Let us consider the algebraic differential equation

$$D(v) - \frac{r D(f)}{1+f} v = 0, \qquad v \in M_D. \tag{1.2}$$

(1.2) has the general solution

$$v = \gamma \exp\left[r \int \frac{D(f)}{1+f}\right]$$

(see [1]), where γ is an arbitrary number. We show that

$$\sum_{\nu=0}^{\infty} \binom{r}{\nu} f^{\nu} \tag{1.3}$$

is also a solution of (1.2). (1.3) converges pointwise for every n, since from $f(1) = 0$ and from the properties of the Dirichlet product follows that for every fixed n, (1.3) has only a finite number of nonzero terms. Moreover, (1.3) can trivially be differentiated term by term. So we have

$$D\Big[\sum_{\nu=0}^{\infty} \binom{r}{\nu} f^{\nu}\Big] = D(f) \sum_{\nu=1}^{\infty} \binom{r}{\nu} \nu f^{\nu-1}. \tag{1.4}$$

By substituting (1.3), (1.4) into (1.2) we have

$$D(f) \sum_{\nu=1}^{\infty} \binom{r}{\nu} \nu f^{\nu-1} - r \frac{D(f)}{1+f} \sum_{\nu=0}^{\infty} \binom{r}{\nu} f^{\nu} =$$

$$= \frac{D(f)}{1+f}\Big[\sum_{\nu=1}^{\infty} \binom{r}{\nu} \nu f^{\nu-1} + \sum_{\nu=1}^{\infty} \binom{r}{\nu} \nu f^{\nu} - r \sum_{\nu=0}^{\infty} \binom{r}{\nu} f^{\nu}\Big] =$$

$$= \frac{D(f)}{1+f}\Big[r + \sum_{\mu=1}^{\infty} \binom{r}{\mu+1} (\mu+1) f^{\mu} + \sum_{\nu=1}^{\infty} \binom{r}{\nu} \nu f^{\nu} - r - r \sum_{\nu=1}^{\infty} \binom{r}{\nu} f^{\nu}\Big] =$$

$$= \frac{D(f)}{1+f} \sum_{\nu=1}^{\infty} \Big[\binom{r}{\nu+1} (\nu+1) + \binom{r}{\nu} \nu - r \binom{r}{\nu}\Big] f^{\nu} = 0.$$

So we obtain that there exists a number $\bar{D}$ such that

$$\bar{D}\exp \int \frac{rD(f)}{1+f} = \sum_{\nu=0}^{\infty} \binom{r}{\nu} f^{\nu}$$

and by substituting $n = 1$ we get $\bar{D} \cdot 1 = 1$, and $\bar{D} = 1$, so the theorem has been proved.

We have shown in paper [1] that

$$1 + f = \exp \int \frac{D(f)}{1+f}, \qquad f \in E, \quad f(1) = 0.$$

Consequently, for arbitrary rational number p, we have

$$(1.5) \qquad (1+f)^p = \exp\left[p \int \frac{D(f)}{1+f}\right] = \sum_{\nu=0}^{\infty} \binom{p}{\nu} f^{\nu}.$$

From elementary properties of the exponential function follows that for rational p, p_1, p_2

$$(1.6) \qquad \begin{aligned} D[(1+f)^p] &= p(1+f)^{p-1} D(f), \\ (1+f)^{p_1}(1+f)^{p_2} &= (1+f)^{p_1+p_2} \end{aligned}$$

holds.

Now we can extend (1.5) for the elements of $\bar{E}$ by the

DEFINITION. Let $h \in \bar{E}$. Then

$$(1.7) \qquad \begin{aligned} h^r :&= (h(1) + h - h(1))^r = (h(1))^r \left(1 + \frac{h - h(1)}{h(1)}\right)^r = \\ &= (h(1))^r \sum_{\nu=0}^{\infty} \binom{r}{\nu} \left(\frac{h - h(1)}{h(1)}\right)^{\nu} \in \bar{E} \end{aligned}$$

for every real r. The definition is correct since it can easily be seen that (1.6) remains true if we replace the numbers p, p_1, p_2 by irrational r, r_1, r_2 (see also Mikusiński [2], page 183).

§ 2. Application to Bernoulli equation

Let us consider the algebraic differential equation of Bernoulli

$$(2.1) \qquad D(x) + ax + bx^m = 0, \qquad a, b \in E, \quad x \in \bar{E}$$

for arbitrary irrational m. If $x \in \bar{E}$, then by the substitution $z = x^{1-m}$ we have $D(z) = (1-m)x^{-m}D(x)$ and (2.1) can be reduced to the linear equation

$$D(z) - (m-1)az = (m-1)b, \qquad z \in M_D, \tag{2.2}$$

so we can apply the results of paper [1] where m was an integer. If we restrict ourselves to $z \in \bar{E}$, then by

$$x = z^{\frac{1}{1-m}}$$

follows that (2.1), (2.2) are equivalent.

In [1] we defined the δ operators for arbitrary $\alpha > 0$, as follows:
if α is irrational, then $\delta(\alpha) = 0$,
if α is integer, then $\delta(\alpha) \in E$, having the value 1 for $n = \alpha$ and zero for $n \neq \alpha$;
if α is rational ($\alpha = \frac{M}{N}$, where M, N are relatively primes), then $\delta(\alpha) = \frac{\delta(M)}{\delta(N)} \notin E$.

We use the following extension of Lemma 2 of paper [1].

LEMMA. (2.1) *has formal solutions in* $\bar{E}$ *if and only if*

$e^{-(m-1)a(1)}$ *is not an integer,*

or

$e^{-(m-1)a(1)}$ *is an integer and*

$$G_m(e^{-(m-1)a(1)}) = 0, \quad \textit{where} \tag{2.2'}$$
$$G_m = (m-1)b \exp\left[-\int (m-1)(a-a(1))\right].$$

The general formal solution of (2.1) *is of the form*

$$x = \left[c\delta\left(e^{-(m-1)a(1)}\right) - \left\{\frac{G_m(n)}{\log n + (m-1)a(1)}\right\}^{\frac{1}{1-m}}\right] e^{-\int (a-a(1))}, \tag{2.3}$$

where in the case that $e^{-(m-1)a(1)}$ *is an integer*

$$\frac{G_m(e^{-(m-1)a(1)})}{\log e^{-(m-1)a(1)} + (m-1)a(1)}$$

denotes the number zero (c is an arbitrary real number).

If (2.3) formally exists, then — as it can easily be seen — (2.3) is a proper solution of (2.1) if and only if

$$C\delta(e^{-(m-1)a(1)}) - \left\{\frac{G_m(n)}{\log n + (m-1)a(1)}\right\} \in \bar{E}. \tag{2.4}$$

I. Let $\gamma = e^{-(m-1)a(1)}$ be not an integer. Then (2.3) formally exists.

If γ is irrational, then $\delta(\gamma) = 0$, if γ is rational, then we must choose $c = 0$, since for $c \neq 0$ $c\delta(\gamma) \notin E$ and (2.4) does not belong to $\bar{E}$. So (2.4) reduces to

$$-\left\{\frac{G_m(n)}{\log n + (m-1)a(1)}\right\}. \tag{2.5}$$

By (2.2′) this function has the value $-\frac{b(1)}{a(1)}$ for $n = 1$. Consequently, (2.5) belongs to $\bar{E}$ iff

$$\frac{b(1)}{a(1)} < 0.$$

II. Let γ be an integer.

If $\gamma = 1$, i.e. $a(1) = 0$, then by the Lemma we have that the formal solution (2.3) exists only for $G_m(1) = (m-1)b(1) = 0$. So $b(1) = 0$ must hold.

Taking again into account the above Lemma it can be seen that (2.4) belongs to $\bar{E}$ iff $c > 0$. If $\gamma > 1$, then by (2.2′) it follows that (2.4) belongs to $\bar{E}$ iff $\frac{b(1)}{a(1)} < 0$. So the following theorem is valid:

THEOREM 2. *Let us consider the Bernoulli equation* (2.1) *for arbitrary (positive or negative) irrational* m, *and let*

$$\gamma = e^{-(m-1)a(1)}.$$

I. *Let* γ *be not an integer. Then* (2.1) *has a solution in* $\bar{E}$ *iff* $\frac{b(1)}{a(1)} < 0$. *For* $\frac{b(1)}{a(1)} < 0$ (2.1) *has exactly one solution in* $\bar{E}$ *of the form*

$$x = \left[-\left\{\frac{G_m(n)}{\log n + (m-1)a(1)}\right\}\right]^{\frac{1}{1-m}} \exp\left[-\int (a - a(1))\right].$$

II. *Let* γ *be an integer.*

If $\gamma = 1$, *i.e.* $a(1) = 0$, *then* (2.1) *has a solution in* $\bar{E}$ *iff* $b(1) = 0$. *If* $b(1) = 0$ *holds, then* (2.1) *has infinitely many solutions in* $\bar{E}$. *The general solution is of the form*

$$\left[c - \left\{\frac{G_m(n)}{\log n}\right\}\right]^{\frac{1}{1-m}} \exp\left[-\int (a - a(1))\right], \tag{2.6}$$

where $c > 0$.

If $\gamma > 1$, *then* (2.1) *has a solution in* $\bar{E}$ *iff*

$$G_m(\gamma) = 0 \quad \textit{and} \quad \frac{b(1)}{a(1)} < 0. \tag{**}$$

If (**) *is satisfied, then* (2.1) *has infinitely many solutions in* $\bar{E}$. *The general solution is of the form* (2.6), *where* c *is an arbitrary real number.*

REFERENCES

[1] FÉNYES, T. and KOSIK, P., The algebraic derivative and integral in the discrete operational calculus, II, *Studia Sci. Math. Hungar.* **10** (1975), 365–380. *MR* **81b**: 44017

[2] MIKUSIŃSKI, J., *Operational calculus*, Vol. 1, 2nd edition, International Series of Monographs in Pure and Applied Mathematics, Vol. 109, Pergamon Press, Oxford–Elmsford, N. Y.; PWN–Polish Scientific Publishers, Warszawa, 1983. *MR* **86b**: 44017

(Received June 13, 1989)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 237–242

ON A CONJECTURE OF KÁTAI

K. KOVÁCS

Let f, f_i denote total additive arithmetical functions. Kátai conjectured that

$$(*) \qquad \sum_{i=1}^{s} f_i(a_i n + b_i) = o(\log n) \quad (a_i \in \mathbb{N},\ b_i \in \mathbb{Z})$$

implies $f_i(p) = c_i \log p$ for all but finitely many primes p.

Conjecture $(*)$ would be a generalization of an important theorem due to E. Wirsing [3], namely that $f(n+1) - f(n) = o(\log n)$ implies $f = c \log$. Kátai proved the conjecture in some special cases [2]:

THEOREM 1. $(*)$ *is valid if* $a_i = 1$ *and* $f_i = c_i f$ $(c_i \in \mathbb{R})$.

We got the following result:

THEOREM 2. *If conjecture* $(*)$ *is true for the choices* $f_i = c_i f$ *and for all* $a_i \in \mathbb{N}$, *then it is true in general.*

Here we show some examples that the choice $a_i = 1$ and $f_i = c_i f$ is not necessary in $(*)$.

THEOREM 3. *Let* f_1 *and* f_2 *be total additive functions and* $A > 0$, $C > 0$, B, D *integers.*

(i) *If for* $B \neq AD$

$$f_1(An + B) + f_2(n + D) = o(\log n),$$

then

$$f_1(n) = -f_2(n) = -c \log n.$$

(ii) *Let* $\Delta = AC(A+1)(C+1)(AD - BC) \neq 0$. *If*

$$f_1(An + B) + f_2(Cn + D) \ \to \ c,$$

1991 *Mathematics Subject Classifications.* Primary 11A25.
Key words and phrases. Characterization of additive functions, the function $c \log n$.

Supported by the Hungarian National Foundation for Scientific Research Grant No. 1901.

Akadémiai Kiadó, Budapest

then

$$f_1(n) = -f_2(n) = c_1 \log n \quad \textit{for all} \quad (n, \Delta) = 1.$$

THEOREM 4. *Let f and g be total additive functions, $A > 0$, $B \neq 0$ integers and $t \in \mathbf{R}$. If*

$$f(An + B) + f(n + B) + tg(n) = o(\log n),$$

then

$$f(n) = c \log n \quad \textit{and for} \quad t \neq 0 \qquad g(n) = (-2c/t) \log n.$$

THEOREM 5. *Let f_1 and f_2 be total additive functions, $D > 0$, $A \neq 0$, B integers, for which $AB > 0$ and $\Delta = ABD(A^2 - B^2) \neq 0$. If*

$$f_1(Dn + A) + f_1(Dn + B) + f_2(n) \to c,$$

then

$$f_1(n) = -f_2(n)/2 = c_1 \log n \quad \textit{for all} \quad (n, \Delta) = 1.$$

THEOREM 6. *Let $f_i = c_i f$ and g be total additive functions and $b_i \in \mathbf{Z}$. If*

$$\sum_{i=1}^{m} f_i(n - 2b_i^2) + g(2n - 1) = o(\log n),$$

then

$$f(n) = c \log n \quad (\textit{and} \quad g(n) = c' \log n \quad \textit{for } (n, 2) = 1).$$

THEOREM 7. *Let f_i $i \in \{1, 2, 3\}$ be total additive functions and $f_{3j} = c_j f_3$ $j \in \{1, \ldots, m\}$.*

(i) *If $a_i \equiv a_2 \equiv a_{3j} \bmod 4$ or*

(ii) *if there exists an odd prime p, for which*

$$a_1 \equiv a_{3j} \bmod 2p \quad \textit{and} \quad a_1 \equiv a_2 \bmod p,$$

then

$$F(n) = f_1(n + a_1) + f_2(n + a_2) + \sum_{j=1}^{m} f_{3j}(n + a_{3j}) = o(\log n)$$

in the case $F \not\equiv 0$ implies

$$f_i(n) = c_i \log n \quad \textit{with} \quad C_1 + C_2 + \sum_{j=1}^{m} c_j C_3 = 0.$$

PROOFS. We need a theorem of Elliott [1].

THEOREM 8. *Let f be an additive function, $A>0$, $C>0$, B, D integers and $\Delta = AC(AD-BC)\neq 0$. If*

$$f(An+B)-f(Cn+D)\ \to\ c,$$

then

$$f(n)=c_1\log n \quad \textit{for all}\ \ (n,\Delta)=1.$$

PROOF OF THEOREM 3. We need the following

LEMMA 1. *Let f be a total additive function, $A>0$, B and D integers. If*

$$f(An+B)-f(n+D)=o(\log n), \tag{1}$$

then $f(n)=c\log n$.

Replacing n by $n-D$ in (1)

$$f(an+b)-f(n)=o(\log n) \tag{2}$$

with $a=A$ and $b=B-AD$. Replacing n by $|b|n$ in (2)

$$f(an+\operatorname{sgn} b)-f(n)=o(\log n). \tag{3}$$

If $b<0$ then (3) gives

$$f(an-1)-f(n)=o(\log n) \tag{4}$$

and replacing n by an^2 in (3) we get

$$f(an+1)+f(an-1)-2f(n)=o(\log n). \tag{5}$$

The difference (5) − (4) implies

$$f(an+1)-f(n)=o(\log n), \tag{6}$$

which we have direct in the case $b>0$ (see (3)).

Using (6) and $(an+1)(an+k)=a[an^2+(k+1)n]+k$ by induction we get

$$f(an+t)-f(n)=o_t(\log n)$$

for all $t\in\mathbb{N}$. The special choice $t=a$ gives

$$f(n+1)-f(n)=o(\log n),$$

so by the mentioned result of Wirsing $f=c\log$. □

(i) Replacing n by $n-D$ we have

$$f_1(An+B)+f_2(n)=o(\log n). \tag{7}$$

By $n \to (A+1)n+B$

$$f_1(An+B)+f_2((A+1)n+B)=o(\log n). \tag{8}$$

So the difference (8) – (7) gives

$$f_2((A+1)n+B)-f_2(n)=o(\log n).$$

Using the lemma we get $f_2(n)=c\log n$ and for the total additive function $f_3=f_1-c\log f_3(an+b)=o(\log n)$. Some elementary methods give $f_3\equiv 0$.

(ii) Replacing n by $(A+1)n+B$ in

$$f_1(An+B)+f_2(Cn+D)\ \to\ c \tag{9}$$

we have

$$f_1(An+B)+f_2(C(A+1)n+CB+D)\ \to\ c'. \tag{10}$$

For the difference (10) – (9) we can apply Theorem 8. □

PROOF OF THEOREM 4. Replacing n by $|B|n$

$$f(An+\operatorname{sgn} B)+f(n+\operatorname{sgn} B)+tg(n)=o(\log n). \tag{11}$$

$n \to An$ in (11) gives

$$f(A^2n+\operatorname{sgn} B)+f(An+\operatorname{sgn} B)+tg(n)=o(\log n). \tag{12}$$

For the difference (12) – (11) we can apply Lemma 1. $f(n)=c\log n$ in (11) implies $g(n)=(-2c/t)\log n$, if $t\neq 0$. □

PROOF OF THEOREM 5. Replacing n by $|A|n$ resp. $|B|n$ and composing the two rows Theorem 8 is applicable. □

PROOF OF THEOREM 6. Replacing n by $n+1$ resp. $2n^2$ in

$$\sum_{i=1}^{m} f_i(n-2b_i^2)+g(2n-1)=o(\log n) \tag{13}$$

we have

$$\sum_{i=1}^{m} f_i(n+1-2b_i^2)+g(2n+1)=o(\log n) \tag{14}$$

and

$$\sum_{i=1} f_i(2n^2-2b_i^2)+g(4n^2-1)=o(\log n). \tag{15}$$

The composition (15) – (14) – (13) contains the f_i's only. Theorem 1 is applicable. □

PROOF OF THEOREM 7. (i) We may assume $A \neq 0$, $A_j \neq A$, $A_j \neq 0$ and $A_j \neq A_s$ in the form after replacing n by $n - a_1$

$$f_1(n) + f_2(n+4A) + \sum_{j=1}^{m} f_{3j}(n+4A_j) = o(\log n). \tag{16}$$

By $n \to 2n + 4A$

$$f_1(n+2A) + f_2(n+4A) + \sum_{j=1}^{m} f_{3j}(n+2A_j+2A) = o(\log n). \tag{17}$$

In the difference (17) – (16)

$$f_1(n+2A) - f_1(n) + \sum_{j=1}^{m} [f_{3j}(n+2A+2A_j) - f_{3j}(n+4A_j)] = o(\log n) \tag{18}$$

let us replace n by $2n$ resp. $2n+2A$ and summarize these two rows. We get

$$\begin{aligned} f_1(n+2A) - f_1(n) + \sum_{j=1}^{m} [f_{3j}(n+A+A_j) + f_{3j}(n+2A+A_j) - \\ - f_{3j}(n+2A_j) - f_{3j}(n+A+2A_j)] = o(\log n). \end{aligned} \tag{19}$$

For the difference (19) – (18) we can apply Theorem 1, i.e. $f_3(n) = c_3 \log n$. Putting this in (18), by Lemma 1, $f_1(n) = c_1 \log n$. (17) gives $f_2(n + a_1) = c' \log n + o(\log n)$, which implies $f_2(n) = c_2 \log n$.

(ii) The proof is similar as in (i). Replacing n by $n - a_1$ we have

$$f_1(n) + f_2(n+pt_1) + \sum_{j=1}^{m} f_{3j}(n+2pt_2) = o(\log n). \tag{20}$$

Replacing n by $pn + (p^2 - p)t$

$$f_1(n+(p-1)t_1) + f_2(n+pt_1) + \sum_{j=1}^{m} f_{3j}(n+(p-1)t_1+2t_2) = o(\log n). \tag{21}$$

In the difference (21) – (20) let us replace n by $2n$, resp. $2n + (p-1)t_1$ and summarize these two rows. Composing it by (20) Theorem 1 is applicable. □

REFERENCES

[1] ELLIOTT, P. D. T. A., *Arithmetic functions and integer products*, Grundlehren der mathematischen Wissenschaften, Bd. 272, Springer-Verlag, New York–Berlin, 1985. *MR* **86j**:11095

[2] KÁTAI, I., Characterization of $\log n$, *Studies in Pure Mathematics*, Birkhäuser-Verlag, Basel–Boston, Mass., 1983, 415–421. *MR* **86m**:11073

[3] WIRSING, E., Additive and completely additive functions with restricted growth, *Recent progress in analytic number theory* (Proc. Sympos., Durham, 1979), Vol. 2, Academic Press, London–New York, 1981, 231–280. *MR* **83a**:10096

(Received July 2, 1990)

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
ALGEBRA ÉS SZÁMELMÉLET TANSZÉK
MÚZEUM KRT. 6–8
H–1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 243–247

CHARACTERIZATIONS OF CAUCHY, NORMAL, AND UNIFORM DISTRIBUTIONS

G. G. HAMEDANI

In a recent paper, Glänzel, Telcs and Schubert [2] established, among other things, an interesting characterization theorem for non-negative continuous random variables based on a simple relation between two truncated moments. In a more recent paper, Glänzel [1] extended that result for arbitrary continuous real-valued random variables (Theorem G) and then applied his theorem to give a very nice characterization of the normal distribution (Proposition G).

Following Glänzel's ideas we first state a slightly different version of Proposition G (Proposition G1) and then apply Theorem G to give characterizations of uniform and Cauchy distributions.

THEOREM G. *Let $(\Omega, \mathcal{A}, \mathcal{P})$ be a given probability space and let $H = [a, b]$ be an interval for some $a < b$ ($a = -\infty$ and $b = +\infty$ might as well allowed). Let $X : \Omega \to H$ be a continuous random variable with the distribution function F and let g and h be two real functions defined on H such that*

$$E\{g(X)|X \geqq x\} = E\{h(X)|X \geqq x\}\lambda_h^g(x), \quad x \in H$$

is defined with some real function λ_h^g. Assume that $g, h \in C^1(H)$, $\lambda_h^g \in C^2(H)$ and F is twice continuously differentiable and strictly monotone function on the set H. Finally, assume that the equation $h\lambda_h^g = g$ has no solution in the int H. *Then F is uniquely determined by the functions g, h and λ_h^g, particularly*

$$F(x) = \int_a^x C \left| \frac{\lambda'(u)}{\lambda(u)h(u) - g(u)} \right| \exp(-s(u))du,$$

where the function s is a solution of the differential equation $s' = \frac{\lambda' h}{\lambda h - g}$ and C is a constant, chosen to make $\int_H dF = 1$.

PROPOSITION G. *Let $X : \Omega \to \mathbf{R}$ be a continuous random variable and let*

$$g(x) = x^2 - mx - \sigma^2, \quad x \in \mathbf{R}$$

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 62E10, 62H05.
Key words and phrases. Characterization of distributions.

Akadémiai Kiadó, Budapest

and

$$h(x) = x - m, \quad x \in \mathbf{R}$$

be two real valued functions with the parameters $m \in \mathbf{R}$ *and* $\sigma \in \mathbf{R}^+$. *The distribution of the random variable* X *is normal if and only if the function* λ_h^g *defined in Theorem G has the form*

$$\lambda_h^g(x) = x, \quad x \in \mathbf{R}.$$

We now restate Proposition G in the following manner:

PROPOSITION G1. *Let* $X : \Omega \to \mathbf{R}$ *be a continuous random variable and let* $h(x) = x - m$, $x \in \mathbf{R}$, *where* $m \in \mathbf{R}$ *is a constant. The distribution of* X *is* $N(m, \sigma^2)$ *if and only if there exist functions* g *and* λ_h^g *defined in Theorem G satisfying the differential equation*

$$\frac{\lambda'(x)}{\lambda(x)h(x) - g(x)} = \frac{1}{\sigma^2}, \quad x \in \mathbf{R}. \tag{1}$$

REMARKS 1. (i) The general solution of the differential equation (1) is

$$\lambda_h^g(x) = \left\{\exp\left[\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right]\right\}\left\{-\frac{1}{\sigma^2}\int\left(\exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right]\right)g(x)dx + D\right\}, \tag{2}$$

where D is a constant.

(ii) In view of (i) one set of functions satisfying both the hypotheses of Theorem G and equation (2) (with $D = 0$) is

$$g(x) = x^2 - mx - \sigma^2, \quad x \in \mathbf{R}$$

$$\lambda_h^g(x) = x, \quad x \in \mathbf{R}.$$

These are the functions considered in Proposition G.

(iii) We note that there are other sets of functions which satisfy both, the hypotheses of Theorem G and equation (2), e. g.

$$g(x) = x^2 - m^2 - \sigma^2, \quad x \in \mathbf{R}$$

$$\lambda_h^g(x) = x + m, \quad x \in \mathbf{R}.$$

The following proposition gives a characterization of the uniform distribution over a bounded interval which we assume, without loss of generality, to be $H = [0, 1]$.

PROPOSITION 1. *Let $X: \Omega \to H$ be a continuous random variable and let $h(x) \equiv 1$, $x \in H$. The distribution of X is uniform over H if and only if there exist functions g and λ_h^g defined in Theorem* G *satisfying the differential equation*

$$\frac{\lambda'(x)}{\lambda(x) - g(x)} = \frac{1}{1-x}, \quad x \in [0,1). \tag{3}$$

PROOF. Let X be $U[0,1]$ and let (for example)

$$g(x) = x, \quad x \in [0,1]$$

$$\lambda_h^g(x) = \frac{1}{2}(1+x), \quad x \in [0,1].$$

Then clearly F, h, g and λ_h^g satisfy both the hypotheses of Theorem G and the equation (3).

Conversely, if there exist functions g and λ_h^g with the stated properties, then $s(x) = \ln(1-x)^{-1}$, $x \in [0,1)$. Thus from Theorem G, X is $U[0,1]$.

REMARKS 2. (i) The general solution of the differential equation (3) is

$$\lambda_h^g(x) = \frac{1}{1-x}\left[-\int g(x)dx + D\right], \quad x \in [0,1). \tag{4}$$

(ii) Clearly, there are other sets of functions which satisfy both the hypotheses of Theorem G and equation (3), e. g. $g(x) = -xe^x$ and $\lambda_h^g(x) = -e^x$ which we obtain from (4) with $D = 0$.

The following set is due to our colleague M. Ahsanullah: $h(x) = 1$, $g(x) = c(x-1)$, c a constant, $0 < |c| < \infty$, and $\lambda_h^g(x) = \frac{c}{2}(x-1)$.

(iii) It may be possible to have similar characterization in which h is not necessarily a constant function (cf. Glänzel [3]).

Finally we like to give, as another application of Theorem G, a characterization of Cauchy distribution which we assume, without loss of generality, to have probability density function (*p df*)

$$f(x) = \left(\pi(1+x^2)\right)^{-1}, \quad x \in \mathbf{R}. \tag{5}$$

The next proposition has the same format as that of Proposition G.

PROPOSITION 2. *Let $X: \Omega \to \mathbf{R}$ be a continuous random variable and let*

$$g(x) \equiv 1, \quad x \in \mathbf{R}$$

and

$$h(x) = x/\left(1+x^2\right)^{1/2}, \quad x \in \mathbf{R}.$$

The p df of X is (5) *if and only if the function* λ_h^g *defined in Theorem G has the form*

$$\lambda_h^g(x) = \left(1+x^2\right)^{1/2}\left(\frac{\pi}{2} - \arctan x\right), \quad x \in \mathbf{R}. \tag{6}$$

PROOF. If X has *p df* (5), then

$$(1-F(x))\,E\{h(X)\,|X \geqq x\} = \frac{1}{\pi}\left(1+x^2\right)^{-1/2}, \quad x \in \mathbf{R},$$

and

$$(1-F(x))\,E\{g(X)\,|X \geqq x\} = \frac{1}{\pi}\left(\frac{\pi}{2} - \arctan x\right), \quad x \in \mathbf{R}.$$

From the definition we obtain that

$$\lambda_h^g(x) = \frac{E\{g(X)\,|X \geqq x\}}{E\{h(X)\,|X \geqq x\}} = \left(1+x^2\right)^{1/2}\left(\frac{\pi}{2} - \arctan x\right), \text{ for all } x \in \mathbf{R}.$$

Now we need only to show that the equation $h\lambda_h^g = g$ has no solution in $\mathbf{R}$, i.e.

$$x\left(\frac{\pi}{2} - \arctan x\right) = 1 \tag{7}$$

has no real solution. Let

$$\varphi(x) = x\left(\frac{\pi}{2} - \arctan x\right), \quad x \in \mathbf{R}.$$

It can be shown that $\varphi'(x) > 0$ for all $x \in \mathbf{R}$ and therefore $\varphi(x)$ is strictly increasing with $\lim_{x\to\infty} \varphi(x) = 1$. Thus (7) has no solution in $\mathbf{R}$.

Conversely, assume that λ_h^g has the form (6). Then

$$s'(x) = \frac{\lambda'(x)h(x)}{\lambda(x)h(x) - g(x)} = \frac{x}{1+x^2}, \quad x \in \mathbf{R},$$

and hence

$$s(x) = \ln\left(1+x^2\right)^{1/2}, \quad x \in \mathbf{R}.$$

We also observe that

$$\frac{\lambda'(x)}{\lambda(x)h(x) - g(x)} = \frac{1}{(1+x^2)^{1/2}}, \quad x \in \mathbf{R}.$$

Thus, from Theorem G, X has a Cauchy distribution with *p df* (5).

REMARKS 3. (i) Proposition 2 also holds if we replace h and λ_h^g by

$$h(x) = \frac{x}{2(1+x^2)^{1/4}}, \quad x \in \mathbf{R},$$

and

$$\lambda_h^g(x) = (1+x^2)^{1/4}\left(\frac{\pi}{2} - \arctan x\right), \quad x \in \mathbf{R}.$$

In this case the equation $h\lambda_h^g = g$ is

$$\frac{x}{2}\left(\frac{\pi}{2} - \arctan x\right) = 1 \tag{8}$$

and if we let

$$\psi(x) = \frac{x}{2}\left(\frac{\pi}{2} - \arctan x\right), \quad x \in \mathbf{R},$$

we find that $\psi(x)$ is strictly increasing on $\mathbf{R}$ with $\lim_{x\to\infty} \psi(x) = \frac{1}{2}$. Thus (8) has no solution in $\mathbf{R}$.

(ii) Here again it may be possible to have similar characterization in which g is not necessarily a constant function. This is indeed the case as was shown by Glänzel [3]. His characterization employs

$$h(x) = -2x/(1+x^2),\ g(x) = (1-x^2)/(1+x^2), \text{ and } \lambda_h^g(x) = x \text{ for all } x \in \mathbf{R}.$$

ACKNOWLEDGEMENT. I am deeply grateful to Professer Glänzel for his careful reading of the manuscript and for calling my attention to his interesting example (Glänzel [3]) which has now been added in Remarks 3 (ii).

REFERENCES

[1] GLÄNZEL, W., A characterization of the normal distribution, *Studia Sci. Math. Hungar.* **23** (1988), 89–91. *MR* **89j**:62026

[2] GLÄNZEL, W., TELCS, A. and SCHUBERT, A., Characterization by truncated moments and its application to Pearson-type distributions, *Z. Wahrsch. Verw. Gebiete* **66** (1984), 173–183. *MR* **86b**:62022

[3] GLÄNZEL, W., A characterization theorem based on truncated moments and its application to some distribution families, *Mathematical statistics and probability theory* (Bad Tatzmannsdorf, 1986), Vol. B, Reidel, Dordrecht–Boston, MA-London, 1987, 75–84. *MR* **89d**:62013

(Received September 20, 1989)

DEPARTMENT OF MATHEMATICS, STATISTICS AND
COMPUTER SCIENCE
MARQUETTE UNIVERSITY
MILWAUKEE, WI 53233
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 249-259

ON THE CONVERGENCE OF THE FOURIER SERIES OF L-ALMOST PERIODIC FUNCTIONS

KÁLMÁN I. KOVÁCS

Abstract

This paper generalizes previous theorems of H. Bohr and Sz. Gy. Révész.

Bohr proved that if f is a limit periodic function satisfying a Lipschitz condition then we can define an arrangement of the Fourier series of f so that there exists a subsequence of partial sums uniformly convergent to the function f. Révész showed that for any continuous periodic function there exists a rearrangement of its Fourier series so that a subsequence of the partial sums tends to f uniformly. Later Révész gave a common generalization of the theorems above: For any uniformly almost periodic function f we can find an ordering of the Fourier series of f such that there exists a subsequence of the partial sums which is uniformly convergent to f. We work out another generalization of the theorems: Let f be a bounded, finite-dimensional L-almost periodic function. Then for any Fourier series of f there exists an ordering of the Fourier series, so that there exists a subsequence of partial sums which is locally uniformly convergent to the function f. Moreover if f is uniformly almost periodic then the convergence is uniform.

Introduction

This paper generalizes previous theorems of Harald Bohr [1, p. 46] and Szilárd Révész [6, Theorem 1]. Bohr proved that if f is a limit periodic function (that is, f can be represented as the uniform limit of a sequence of purely periodic continuous functions [1, p. 35]) and satisfies a Lipschitz condition

$$\sup_{x\in\mathbb{R}} |f(x+\delta) - f(x)| < c\delta^{\varrho} \qquad (\delta > 0)$$

with some constants $c > 0$, $0 < \varrho \leqq 1$, then we can define an arrangement of the Fourier series of f such that there exists a subsequence of partial sums uniformly convergent to the function f.

Révész showed that for any continuous periodic function there exists a rearrangement of its Fourier series so that a subsequence of the partial sums tends to f uniformly. In [5, Theorem 2] Révész gave a common generalization of the theorems above: For any uniformly almost periodic (u.a.p.) function f we can find an ordering of the Fourier series of f such that there exists a subsequence of the partial sums which is uniformly convergent to f.

1980 *Mathematics Subject Classifications.* Primary 42A75; Secondary 42A20, 42C20.

Key words and phrases. Fourier series of u.a.p. and L.a.p. functions, de la Vallée Poussin means, Bochner–Fejér summation, locally uniform convergence of partial sums.

Akadémiai Kiadó, Budapest

We work out another generalization of the theorems mentioned above extending the result to bounded, finite-dimensional L-almost periodic (L.a.p.) functions.

§ 1

Let us recall Levitan's definition of L.a.p. functions [4, p. 143].

DEFINITION 1. A continuous function f is L.a.p. if there exists a sequence of real numbers $\{\lambda_n\} = \Lambda$ (depending on f) such that for every $\varepsilon > 0$ and $N \in \mathbf{N}$ there exist an integer $m = m(\varepsilon, N)$ and a real $\eta = \eta(\varepsilon, N)$ with the following property:

For every $t \in \mathbf{R}$ satisfying the system of inequalities

$$|\lambda_n \cdot t| < \eta \pmod{2\pi} \qquad (n = 1, \dots, m) \tag{1}$$

we also have

$$|f(x+t) - f(x)| < \varepsilon \qquad (|x| < N). \tag{2}$$

DEFINITION 2. An L.a.p. function is called finite-dimensional if the set $\mathcal{M}_f$ depending only on the function f (see (4)) is a finite-dimensional vector space over the rational field $\mathbf{Q}$.

THEOREM. *Let F be a bounded, finite-dimensional L.a.p. function with Fourier series* (5). *Then there exists an ordering of the Fourier series (i.e. an ordering of $\mathcal{M}_f$), such that there exists a subsequence of partial sums which is locally uniformly convergent to the function f. Moreover, if f is a u.a.p. function then the convergence is uniform.*

§ 2. Fourier analysis of L.a.p. functions

2.1. Following Levitan [4] and Levin [3, pp. 73–74], we shall consider functions that satisfy

$$\lim_{T\to\infty} \left(\inf_{\tau} \frac{1}{2T} \int_{\tau-T}^{\tau+T} |f(t)|dt \right) = K < \infty. \tag{3}$$

Let the set $\Lambda = \{\lambda_n : n \in \mathbf{N}\}$ be a generator-sequence of f as described in Definition 1 and let $\mathcal{M} = \mathcal{M}_f$ denote the vector space generated by Λ over $\mathbf{Q}$, that is

$$\mathcal{M} = \left\{ \sum_{j=1}^{k} x_j \lambda_j : k \in \mathbf{N}, x_j \in \mathbf{Q}, \lambda_j \in \Lambda \right\}. \tag{4}$$

Note that $\mathcal{M}$ is uniquely defined[1] for f, i.e. $\mathcal{M}$ is independent of Λ (see [3, Theorem 8, p. 75]).

Let the sequence $\{\beta_n\}$ denote a basis of Λ over $\mathbf{Q}$. We can define the Fourier series

$$f \sim \sum_{M \in \mathcal{M}} \alpha(M) e^{iMx}, \tag{5}$$

where

$$\alpha(M) = \int_{T_f} f(p)\overline{\chi_M(p)}dp \qquad (M \in \mathcal{M}). \tag{6}$$

(Here $\chi_M(p)$ is the character of the topological group T_f extending e^{iMx} $(x \in \mathbf{R})$ to T_f, dp is the Haar measure on T_f, and $f(p)$ is an extension of f to T_f; see [3, (2.35) and (2.36), p. 74]).

If f satisfies (3) then the Fourier series (5) exists.

In particular, if f is a bounded, L.a.p. function then it has Fourier series. Let us note that there exists a bounded L.a.p. function to what Fourier series is not unique [4, pp. 151–153].

From now on we shall use the notation above for all the Fourier series of L.a.p. functions. If we mention Fourier series we think of a formal series (5) with coefficients (6) where $\mathcal{M} = \mathcal{M}_f$, its basis $\{\beta_n\}$ and also the generator sequence Λ are fixed.

2.2. If f is an L.a.p. function satisfying (3) then there exist finite trigonometric sums tending locally uniformly to the function f [3, Theorem 10, p. 77].

For bounded f a sequence of trigonometrical sums tending to f can be expressed from the Fourier series by the Bochner-Fejér summation [4, pp. 158–163].

We can express any $\lambda_l \in \Lambda$ as the linear combination of finitely many basis elements β_j. Therefore for any $m = m(\varepsilon, N) > 0$, $m \in \mathbf{N}$ and any $\eta = \eta(\varepsilon, N) > 0$ there exist integers $d_m = d(\lambda_1, \ldots, \lambda_m)$ and $p_m = p(\lambda_1, \ldots, \lambda_m)$ and a real $\delta_m = \delta(p_m, \lambda_1, \ldots, \lambda_m) > 0$ such that

$$\lambda_l = \sum_{j=1}^{d_m} k_{l,j} \frac{\beta_j}{p_m} \qquad (l = 1, \ldots, m), \quad k_{l,j} \in \mathbf{Z} \tag{7}$$

and if for a $t \in \mathbf{R}$

$$\left| \frac{\beta_j}{p_m} t \right| < \delta_m \pmod{2\pi} \quad (j = 1, \ldots, d_m) \tag{8}$$

[1] Levin [3] denotes by $\mathcal{M}$ the modulus of the exponents of the characters, but here, following Levitan [4], we use the notation for the vector space generated by this modulus.

hold, then t satisfies (1), too.

We introduce the following notations

$$\underline{\beta}_m = \left(\frac{\beta_1}{p_m}, \dots, \frac{\beta_{d_m}}{p_m}\right) \in \mathbf{R}^{d_m},$$

$$\underline{k} = (k_1, \dots, k_{d_m}) \in \mathbf{Z}^{d_m},$$

(9)

$$\langle \underline{k}, \underline{\beta}_m \rangle = \sum_{j=1}^{d_m} k_j \frac{\beta_j}{p_m} \in \mathcal{M},$$

$$\underline{Q} = (Q_1, \dots, Q_{d_m}) \in \mathbf{N}^{d_m}.$$

The well-known Fejér kernel is

$$\mathcal{K}_Q(p) = \sum_{|k|<Q} \left(1 - \frac{|k|}{Q}\right) \chi_k(p) \qquad (p \in T_f),$$

$$\mathcal{K}_Q(x) = \frac{1}{Q} \frac{\sin^2\left(\frac{Qx}{2}\right)}{\sin^2\left(\frac{x}{2}\right)} \qquad (x \in \mathbf{R}),$$

and the Bochner–Fejér kernel is

(10)

$$\mathcal{K}_{\underline{\beta}_m, \underline{Q}}(p) = \prod_{j=1}^{d_m} \mathcal{K}_{Q_j}\left(\frac{\beta_j}{p_m} p\right) = \sum_{|k_j|<Q_j} \left(1 - \frac{|k_1|}{Q_1}\right) \cdots \left(1 - \frac{|k_{d_m}|}{Q_{d_m}}\right) \chi_{\langle \underline{k}, \underline{\beta}_m \rangle}(p)$$

$$(j = 1, \dots, d_m).$$

We form for any f satisfying (3) the Bochner–Fejér polynomial corresponding to $\underline{Q}$ and $\underline{\beta}_m$ as

$$\sigma_{\underline{\beta}_m, \underline{Q}}(x) = \int_{T_f} f(x+p) \mathcal{K}_{\underline{\beta}_m, \underline{Q}}(p) dp =$$

(11)

$$= \sum_{|k_j|<Q_j} \left(1 - \frac{|k_1|}{Q_1}\right) \cdots \left(1 - \frac{|k_{d_m}|}{Q_{d_m}}\right) \alpha(\langle \underline{k}, \underline{\beta}_m \rangle) e^{i\langle \underline{k}, \underline{\beta}_m \rangle x}$$

$$(x \in \mathbf{R},\ j = 1, \dots, d_m)$$

where $\alpha(\langle \underline{k}, \underline{\beta}_m \rangle)$ is defined in (6).

Let every coordinate sequences $\{Q_{n,i}\}$ of the vectors $\underline{Q}_n = (Q_{n,1}, \dots, Q_{n,d_n})$ be increasing.

Applying Theorem 11 of [3, pp. 78–79] for the sequences $\{\varepsilon_n = \frac{1}{n}\}$ and $\{N_n = n\}$ $(n \in \mathbf{N})$ we get increasing sequences $\{d_n\}$ and $\{Q_n^{(0)}\}$ such that for

every $\underline{Q} = (Q_1, \dots, Q_{d_n})$, $Q_j > Q_n^{(0)}$ $(j = 1, \dots, d_n)$

$$\left|\sigma_{\underline{\beta}_n, \underline{Q}}(x) - f(x)\right| < \frac{1}{n} \qquad (|x| < n). \tag{12}$$

Applying (12) to the de la Vallée Poussin polynomials

$$V_{\underline{\beta}_n, Q}(x) =$$

$$= \sum_{\substack{e_j \in \{0,1\} \\ (j=1,\dots,d_n)}} (-1)^{d_n + (e_1 + \dots + e_{d_n})} \cdot 2^{(e_1 + \dots + e_{d_n})} \sigma_{\underline{\beta}_n, \begin{pmatrix} (1+e_1)Q \\ \vdots \\ (1+e_{d_n})Q \end{pmatrix}}(x) \tag{13}$$

we get

$$\left|V_{\underline{\beta}_n, Q}(x) - f(x)\right| < 3^{d_n} \frac{1}{n} \qquad (|x| < n) \tag{14}$$

whenever $Q > Q_n^{(0)}$. Hence we can formulate the following

LEMMA 1. *If f is a bounded, finite-dimensional L.a.p. function then for any increasing sequence $\{Q_n\}$ with $Q_n > Q_n^{(0)}$ we have*

$$\left|V_{\underline{\beta}_n, Q_n}(x) - f(x)\right| < 3^d \frac{1}{n} \qquad (|x| < n). \tag{15}$$

REMARK 1. If f is a u.a.p. function then $V_{\underline{\beta}_n, Q_n}(x)$ tends uniformly to f since in (13) we can write $x \in \mathbf{R}$ instead of $|x| < n$. See [4, p. 161, (3.4.6)].

2.3. B. Ya. Levin [3] proved the Parseval's formula for L.a.p. functions satisfying some conditions. On purpose to prove the Theorem it is enough to state the following weaker

LEMMA 2 („Bessel's inequality"). *Let f be an L.a.p. function satisfying condition* (3). *Then*

$$\sum_{M \in \mathcal{M}} |\alpha(M)|^2 \leqq K.$$

PROOF. Cf. [3, Theorem 6 on p. 73 and Theorem 7 on p. 74]. □

2.4. Next we give an example of a bounded, finite-dimensional L.a.p. function which is not u.a.p.

LEMMA 3. *Let f be an L.a.p. function and $\Lambda = \{\lambda_k\}$ be a generator-sequence of f. Let g be continuous on* Range(f). *Then $(g \circ f)$ is also an L.a.p. function and Λ is a generator-sequence of $(g \circ f)$, too.*

PROOF. Let $\varepsilon > 0$ and $N \in \mathbf{N}$ be arbitrary. Since f is continuous, $f([-N; N]) = [a; b] \subset$ Range(f) and there exists $\delta_0 = \delta_0(N) > 0$ such that g

is uniformly continuous on the interval $I=[a-\delta_0; b+\delta_0]\cap \mathrm{Range}(f)$. Hence for every $\varepsilon>0$ there exists a δ, $\delta<\delta_0$ such that

$$|g(z)-g(y)|<\varepsilon, \quad \text{if} \quad |y-z|<\delta, \quad y \text{ and } z\in I. \tag{16}$$

Let $t\in\mathbf{R}$ be a $\delta-N$ almost period of f, that is

$$|f(x+t)-f(x)|<\delta \qquad (|x|<N). \tag{17}$$

Since $y=f(x)\in[a;b]\subset I$ and $z=f(x+t)\in I$ by (17), applying (16) we get

$$|(g\circ f)(x+t)-(g\circ f)(x)|<\varepsilon \qquad (|x|<N). \tag{18}$$

Hence all the $\delta-N$ almost periods of f are $\varepsilon-N$ almost periods of $(g\circ f)$, too. □

Note that if f is a u.a.p. function and g is uniformly continuous then $(g\circ f)$ is also a u.a.p. function (see [1, p. 3]).

EXAMPLE. Let $f(x)=2-\sin(x)-\sin(\sqrt{2}x)$, $g(x)=\frac{1}{f(x)}$, $h(x)=\min\limits_{k\in\mathbf{Z}}|x-k|$. Then $v(x)=(h\circ g)(x)$ is a bounded, finite-dimensional L.a.p. function which is not u.a.p.

NOTE. A similar function is described in [3, pp. 100–101] without detailed proof.

PROOF. $v(x)$ is obviously bounded.

f is a u.a.p. function, $f\neq 0$ and $\Lambda(f)=\{1;\sqrt{2}\}$ is finite-dimensional. Applying Lemma 3 twice g and also v are L.a.p. functions, and moreover, Λ is a generator-system belonging to g and also to v. Hence v is a finite-dimensional L.a.p. function.

Finally, we have to prove that v is not u.a.p. It is sufficient to show that v is not uniformly continuous. We will prove that for every $\delta>0$ we can find $x', x''\in\mathbf{R}$ such that

$$|x'-x''|<\delta, \quad \text{but} \quad |v(x')-v(x'')|>\frac{1}{4}. \tag{19}$$

To (19) it is enough to show that for some $N\in\mathbf{N}$

$$|g(x')-N|<\frac{1}{8} \quad \text{and} \quad \left|g(x'')-\left(N+\frac{1}{2}\right)\right|<\frac{1}{8}. \tag{20}$$

In other words

$$\begin{aligned} a_1=\frac{1}{N+\frac{1}{8}}<2-\sin x'-\sin\sqrt{2}x'<\frac{1}{N-\frac{1}{8}}=b_1,\\ a_2=\frac{1}{N+\frac{5}{8}}<2-\sin x''-\sin\sqrt{2}x''<\frac{1}{N+\frac{3}{8}}=b_2. \end{aligned} \tag{21}$$

Obviously, $[a_1; b_1] \cap [a_2; b_2] = 0$ and $b_2 < a_1$.

In view of the continuity of f it is sufficient to prove that for every positive δ there exists $N \in \mathbf{N}$ and $x_0 \in \mathbf{R}$ such that

$$2 - \sin x_0 - \sin \sqrt{2} x_0 < \frac{1}{N + \frac{5}{8}}, \tag{22}$$

$$2 - \sin(x_0 + \delta) - \sin \sqrt{2}(x_0 + \delta) > \frac{1}{N - \frac{1}{8}}, \tag{23}$$

hold simultaneously. If we choose N to be large enough (for example $N > 400/\delta^2$) then (23) follows from (22). But for every $N \in \mathbf{N}$ we can find $x_0 \in \mathbf{R}$ which satisfies (22), since it follows from the Kronecker Theorem [2, p. 382] that

$$\inf_{x \in \mathbf{R}} (2 - \sin x - \sin \sqrt{2} x) = 0. \qquad \square$$

§ 3. Proof of the Theorem

Let f be a bounded, finite-dimensional L.a.p. function, let Λ be a finite-dimensional generator-sequence of f and let $\underline{\beta} = (\beta_1, \ldots, \beta_d)$ be a basis belonging to Λ.

3.1. Take the sequences $\varepsilon_n = \frac{1}{n}$ and $N_n = n$ $(n \in \mathbf{N})$. In view of Definition 1, for every $n \in \mathbf{N}$ there exist sequences $m_n = m(n) \in \mathbf{N}$ and $\eta_n = \eta(n) \in \mathbf{R}$ such that $t \in \mathbf{R}$ is $\varepsilon_n - N_n$ almost period of f if t satisfies

$$|\lambda_l t| < \eta_n \pmod{2\pi}, \qquad \lambda_l \in \Lambda \quad (l = 1, \ldots, m_n). \tag{24}$$

Moreover, there are sequences $\{\delta_n\}$, $\delta_n = \delta(n) > 0$ and $\{q_n\}$, $q_n = q(n) \in \mathbf{N}$, q_n increasing such that

$$\lambda_l = \sum_{j=1}^{d} k_{l,j}^{(n)} \frac{\beta_j}{q_n!}, \qquad k_{l,j}^{(n)} \in \mathbf{Z} \quad (l = 1, \ldots, m_n), \tag{25}$$

and if $t \in \mathbf{R}$ satisfies the system of inequalities

$$\left| \frac{\beta_j}{q_n!} t \right| < \delta_n \pmod{2\pi} \qquad (j = 1, \ldots, d) \tag{26}$$

then also (24) holds true.

Introduce the following notations:

$$\underline{\beta}_n = \left(\frac{\beta_1}{q_n!}, \ldots, \frac{\beta_d}{q_n!} \right) \in \mathbf{R}^d,$$

$$\|\underline{k}\| = \max_{j=1,\dots,d} |k_j|,$$

$$S_{\underline{\beta}_n,Q}(x) = \sum_{\|\underline{k}\|<Q} a(\langle \underline{k}, \underline{\beta}_n \rangle) e^{i(\langle \underline{k}, \underline{\beta}_n \rangle)x},$$

$$\mathcal{M}_{\underline{\beta}_n,Q} = \left\{ \langle \underline{k}, \underline{\beta}_n \rangle, \|\underline{k}\| < Q \right\}.$$

3.2. LEMMA 4. *With the notation above there exists a sequence* $\{Q_n\}$, $Q_n \to \infty$ *which satisfies the following conditions for every* $n \in \mathbb{N}$*:*

$$\lambda_l \in \mathcal{M}_{\underline{\beta}_n,Q_n} \qquad (l = 1, \dots, m_n), \tag{27}$$

$$\mathcal{M}_{\underline{\beta}_{n-1},2Q_{n-1}} \subset \mathcal{M}_{\underline{\beta}_n,Q_n}, \tag{28}$$

$$Q_n > Q_n^{(0)}, \tag{29}$$

$$Q_n > 28d, \tag{30}$$

$$\sum_{Q_n \le \|\underline{k}\| < 2Q_n} |a(\langle \underline{k}, \underline{\beta}_n \rangle)|^2 < \frac{\varepsilon_n^2}{d^2 \log Q_n}. \tag{31}$$

PROOF. Choosing Q_n large enough conditions (27), (29), (30) are satisfied trivially. If $Q_n > 2q_n Q_{n-1}$, (28) will be satisfied, too. Let us denote Q_n^* the minimum of Q_n satisfying (27)–(30). Put for fixed $n \in \mathbb{N}$

$$\varphi(t) = \varphi_n(t) = \sum_{\|\underline{k}\|<t} |a(\langle \underline{k}, \underline{\beta}_n \rangle)|^2.$$

In view of Lemma 3 $\varphi(t)$ is integrable on $[1, \infty)$. Hence there exists a sequence $\{Q^{(m)}\}$, $Q^{(m+1)} > 2Q^{(m)}$ such that

$$\int_{Q^{(m)}}^{2Q^{(m)}} \varphi(t)dt = O\left(\frac{1}{\log Q^{(m)}}\right),$$

see, e.g. [5, Lemma 5]. Consequently, for $m > m(n)$ (31) will be satisfied with $Q^{(m)}$ in place of Q_n. Choosing an $m > m(n)$ such that $Q^{(m)} > Q_n^*$, $Q_n = Q^{(m)}$ defines a sequence $\{Q_n\}$ satisfying conditions (27)–(31). □

Note that if $\{Q_n\}$ is a sequence satisfying condition (28) then we also have

$$\lim_{n\to\infty} \mathcal{M}_{\underline{\beta}_n, Q_n} = \mathcal{M}. \tag{32}$$

3.3. Let us introduce on the d-dimensional torus

$$\mathbf{T}^d = \mathbf{R}^d / 2\pi\mathbf{Z}^d$$

the following notations:

$$\underline{x}^{(n)} = \left(\frac{\beta_1}{q_n!}x_1, \ldots, \frac{\beta_d}{q_n!}x_d\right) \in \mathbf{T}^d,$$

$$S_n^*(\underline{x}) = S_{\underline{\beta}_n, Q_n}^*(\underline{x}) = \sum_{\|\underline{k}\|<Q} a(\langle \underline{k}, \underline{\beta}_n \rangle) e^{i\langle \underline{k}, \underline{x}^{(n)} \rangle}.$$

We also put

$$V_n^*(\underline{x}) = V_{\underline{\beta}_n, Q_n}^*(\underline{x}) = \sum_{\|\underline{k}\|<2Q_n} p_{Q_n}(\underline{k}) a(\langle \underline{k}, \underline{\beta}_n \rangle) e^{i\langle \underline{k}, \underline{x}^{(n)} \rangle}, \tag{34}$$

where

$$p_Q(\underline{k}) = \prod_{j=1}^{d} p_Q(k_j),$$

$$p_Q(k) = \begin{cases} 1 & \text{if } |k| < Q, \\ 2 - \frac{|k|}{Q} & \text{if } Q \leqq |k| < 2Q, \\ 0 & \text{otherwise,} \end{cases}$$

that is

$$V_n^*(\underline{x}) = S_n^*(\underline{x}) + \sum_{Q_n \leqq \|\underline{k}\| < 2Q_n} p_{Q_n}(\underline{k}) a(\langle \underline{k}, \underline{\beta}_n \rangle) e^{i\langle \underline{k}, \underline{x}^{(n)} \rangle}. \tag{35}$$

Now we apply [5, Lemma 6] with $\eta = \frac{\varepsilon_n^2}{d^2}$, $b(k) = A(\langle \underline{k}, \underline{\beta}_n \rangle)$ and with Q_n in place of n. The conditions of [5, Lemma 6] are satisfied according to (30) and (31) of Lemma 4. Hence there exists a sequence $\omega_{\underline{k}}^{(n)} \in \{0, 1\}$, $(Q_n \leqq \|\underline{k}\| < 2Q_n)$ such that

$$\sup_{\underline{x}\in\mathbf{T}^d} \left| V_n^*(\underline{x}) - \left(S_n^*(\underline{x}) + \sum_{Q_n \leqq \|\underline{k}\| < 2Q_n} \omega_{\underline{k}}^{(n)} a(\langle \underline{k}, \underline{\beta}_n \rangle) e^{i\langle \underline{k}, \underline{x}^{(n)} \rangle} \right) \right| < 8\varepsilon_n. \tag{36}$$

Let us return to $\mathbf{R}$ using for any $\Psi\colon \mathbf{T}^d \to \mathbf{R}$ the diagonal function

$$\xi(x) = \Psi\Big(\frac{\beta_1}{q_n!}x, \dots, \frac{\beta_d}{q_n!}x\Big).$$

We get for the de la Vallée Poussin polynomials $V_n(x) = V_{\underline{\beta}_n, Q_n}(x)$ that with $S_n(x) = S_{\underline{\beta}_n, Q_n}(x)$ we have

$$\sup_{x\in\mathbf{R}}\Big|V_n(x) - \Big(S_n(x) + \sum_{Q_n \le \|\underline{k}\| < 2Q_n} \omega_{\underline{k}}^{(n)} a(\langle \underline{k}, \underline{\beta}_n\rangle) e^{i\langle \underline{k}, \underline{\beta}_n\rangle x}\Big)\Big| < 8\varepsilon_n. \tag{37}$$

3.4. Let us define the ordering ν in the following way.

$$\nu = \bigcup_{n=1}^{\infty} \nu_n,$$

$$\nu_n\colon \mathcal{M}_{\underline{\beta}_n, 2Q_n} \setminus \mathcal{M}_{\underline{\beta}_{n-1}, 2Q_{n-1}} \leftrightarrow ((4Q_{n-1}+1)^d, (4Q_n+1)^d], \tag{38}$$

$$\nu_n = \begin{cases} \mathcal{H}_n \leftrightarrow & ((4Q_{n-1}+1)^d, (2Q_n+1)^d], \\ \mathcal{J}_n \leftrightarrow & ((2Q_n+1)^d, N_n], \\ \mathcal{K}_n \leftrightarrow & (N_n, (4Q_n+1)^d], \end{cases}$$

where

$$\mathcal{H}_n = \mathcal{M}_{\underline{\beta}_n, Q_n} \setminus \mathcal{M}_{\underline{\beta}_{n-1}, 2Q_{n-1}},$$

$$\mathcal{J}_n = \{M \in \mathcal{M}_{\underline{\beta}_1, 2Q_n} \setminus \mathcal{M}_{\underline{\beta}_n, Q_n} : M = \langle \underline{k}, \underline{\beta}_n\rangle, \omega_{\underline{k}}^{(n)} = 1\},$$

$$\mathcal{K}_n = \{M \in \mathcal{M}_{\underline{\beta}_n, 2Q_n} \setminus \mathcal{M}_{\underline{\beta}_n, Q_n} : M = \langle \underline{k}, \underline{\beta}_n\rangle, \omega_{\underline{k}}^{(n)} = 0\},$$

$$N_n = (2Q_n+1)^d + |\mathcal{J}_n|.$$

That is, ν_n counts first the exponents from $\mathcal{H}_n$, then those exponents from $\mathcal{M}_{\underline{\beta}_n, 2Q_n} \setminus \mathcal{M}_{\underline{\beta}_n, Q_n}$ which have $\omega_{\underline{k}}^{(n)} = 1$, and finally those with $\omega_{\underline{k}}^{(n)} = 0$.

Since $\{Q_n\}$ satisfies (28) and (32), ν defines an ordering on $\mathcal{M}$.

3.5. Consider the ν-arranged Fourier series of f and put for the N-th partial sum ${}_\nu S_N(x)$. According to (38) we have

$${}_\nu S_{N_n}(x) = S_n(x) + \sum_{Q_n \le \|\underline{k}\| < 2Q_n} \omega_{\underline{k}}^{(n)} a(\langle \underline{k}, \underline{\beta}_n\rangle) e^{i\langle \underline{k}, \underline{\beta}_n\rangle x}. \tag{39}$$

Hence we get from (37) that

$$|V_n(x) - {}_\nu S_{N_n}(x)| < 8\varepsilon_n \qquad (x \in \mathbf{R}). \tag{40}$$

Finally, an application of Lemma 1 concludes the proof of the Theorem, also taking into consideration Remark 1 when f is u.a.p.

REFERENCES

[1] Besicovitch, A. S., *Almost periodic functions*, Cambridge University Press, Cambridge, 1932. *Zbl* **4**, 253

[2] Hardy, G. H. and Wright, E. M., *An introduction to the theory of numbers*, Fifth edition, The Clarendon Press, Oxford University Press, Oxford, 1979. *MR* **81i**:10002

[3] Levin, B. Ya., On the almost periodic functions of Levitan, *Ukrain. Mat. Žurnal* **1** (1949), 49–101 (in Russian). *MR* **14**–370

[4] Levitan, B. M., *Počti-periodičeskie funkcii* [Almost periodic functions], Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow, 1953 (in Russian). *MR* **15**–700

[5] Révész, Sz. Gy., On the convergence of Fourier series of u.a.p. functions, *J. Math. Anal. Appl.* (to appear).

[6] Révész, Sz. Gy., Rearrangements of Fourier series, *J. Approx. Theory* **60** (1990), 101–121. *MR* **90m**:42042

(Received October 26, 1989)

BUDAPESTI MŰSZAKI EGYETEM
GÉPÉSZMÉRNÖKI KAR
MATEMATIKA TANSZÉK
EGRI JÓZSEF U. 1
H–1521 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 261–266

ON THE DIRECTIONAL DERIVATIVES

I. JOÓ and M. PALKO

J. Marcinkiewicz's theorem on universal functions has been extended to L_p-spaces in [1] for $0<p\leqq 1$ and the problem has also been raised for $p>1$. The negative answer was given independently by the authors of [2]–[4]. The most elegant solution is due to M. Horváth, who has proved, roughly speaking, that "the difference quotient is bounded below in mean for every function which is non-constant in every direction". The exact meaning of this and the proof is given in M. Horváth's paper [2] (see also the reference there).

The aim of the present note is the investigation of the pointwise properties of partial derivatives of "smooth" functions. The motivation for this is Horváth's result mentioned above.

Let $\Omega \subset \mathbf{R}^N$ $(N \geqq 1)$ be a domain and f be any function $f : \Omega \subset \mathbf{R}^M$ $(M \geqq 1)$. Denote (as usual) $Df(X, e)$ the set of "one-sided directional sequential derivatives of f in the set $x \in \Omega$ in the direction e", i.e.

$$Df(x,e) := \left\{ \lim_{\lambda_n \to +0} \frac{f(x+\lambda_n) - f(x)}{\lambda_n}, \quad \text{if exists} \right\}.$$

The following theorem shows that the compact non-empty set $Df(x, e)$ can be "exotic" even for smooth function f.

THEOREM. (1) *There exists a function $f : \mathbf{R}^2 \to \mathbf{R}^2$ such that f and $f^{-1} \in$ $\in$ Lip1 and*

$$\left\{ q \in \mathbf{R}^2 : \exists p \in Df(x, e_0) \text{ such that } q = \frac{p}{\|p\|} \right\} = S^1$$

holds only for one direction e_0. S^1 denotes the unit circle, i.e. $S^1 := \{q \in \mathbf{R}^2 : \|q\| = 1\}$.

(2) *Let $N \geqq 2$. There exists a function $f : \mathbf{R}^N \to \mathbf{R}^N$ such that f and $f^{-1} \in$ Lip1 and*

$$\left\{ q \in \mathbf{R}^N : \exists p \in Df(x, e) \text{ such that } q = \frac{p}{\|p\|} \right\} = S^{N-1}$$

1991 *Mathematics Subject Classifications.* Primary 26B05; Secondary 26A16.
Key words and phrases. Directional derivative, Lipschitzian functions.

Akadémiai Kiadó, Budapest

holds for every $e \in \mathbf{R}^N \setminus \{0\}$. *Here* $S^{N-1} := \{q \in \mathbf{R}^N : \|q\| = 1\}$.

It would be possible to think that the curve $f(x+te_0)$ is rotated around $f(x)$ for $t \to +0$. But if the rotation changes the direction after any rotation appropriately, then we can approximate $f(x)$ without rotating around it. The numerical details are given below in the proof of (i).

PROOF OF (1). Define the function $f(x,y)$ by the following formulas. Let $f(0,0) = (0,0)$, for $x > 0$ let

$$f(x,0) = \left(x \cos \ln x, (-1)^{\left[\frac{\ln x}{2\pi}\right]} x \cdot \sin \ln x\right)$$

i.e. the rotation of $(x,0)$ by the angle $(-1)^{[\ln x/2\pi]} \ln x$, $[z]$ denotes the entire part of z.

$$f(-x,0) = \left(x \cos\left(\frac{\pi}{2} \cos\left(\frac{\ln x}{2}\right)\right), -x \sin\left(\frac{\pi}{2} \cos\left(\frac{\ln x}{2}\right)\right)\right)$$

the rotation of $(x,0)$ by the angle $-\frac{\pi}{2} \cos \frac{\ln x}{2}$.

It is easy to see that the mod 2π distance of $(-1)^{[\ln x/2\pi]} \ln x$ and $-\frac{\pi}{2} \cos \frac{\ln x}{2}$ is between $\frac{\pi}{2}$ and π. Now let $(x,y) = (r \cos \alpha, r \sin \alpha)$. We define $f(x,y)$ as the rotation of the point $(r,0)$ by an angle which depends linearly on α. Namely define

$$\{z\}_{2\pi} := z - 2k_0\pi, \quad k_0 := \max\{k : 2k\pi \leqq z\}.$$

For $0 \leqq \alpha \leqq \pi$ let $f(x,y)$ be the rotation of $(r,0)$ by the angle

$$\frac{\pi - \alpha}{\pi} \left\{(-1)^{\left[\frac{\ln r}{2\pi}\right]} \ln r\right\}_{2\pi} + \frac{\alpha}{\pi} \left(-\frac{\pi}{2} \cos\left(\frac{\ln r}{2}\right)\right)$$

(for $r = e^{(4k+2)\pi}$ by the angle $\frac{\pi-\alpha}{\pi} 2\pi + \frac{\alpha}{2} = 2\pi - \frac{3\pi}{2}$) for $-\pi \leqq \alpha \leqq 0$ $f(x,y)$ be the rotation of $(r,0)$ by the angle

$$\frac{\pi - \alpha}{|\pi|} \left(\left\{(-1)^{\left[\frac{\ln r}{2\pi}\right]} \ln r\right\}_{2\pi} - 2\pi\right) + \frac{|\alpha|}{\pi} \left(-\frac{\pi}{2} \cos\left(\frac{\ln r}{2}\right)\right)$$

(for $r = e^{(4k+2)\pi}$ by the angle $-\frac{\alpha}{2}$).

It is easy to see that for $e_0 = (1,0)$, $Df((0,0),e_0) = S^1$ and for any $e \neq e_0$, $|e| = 1$, $Df((0,0),e)$ is only a part of S^1 (does not contain e_0). It remains to verify the Lipschitz property of f and f^{-1}. Suppose indirectly that there exist two sequences x_n, $y_n \in \mathbf{R}^2$ such that

(a) $$\left|\frac{f(x_n) - f(y_n)}{x_n - y_n}\right| \to \infty$$

or

$$\text{(b)} \qquad \left|\frac{f(x_n)-f(y_n)}{x_n-y_n}\right| \to 0.$$

Since $|x_n|=|f(x_n)|$, $|y_n|=|f(y_n)|$ we must have in both cases $|\frac{x_n}{y_n}| \to 1$. Denote α_n, β_n, $\alpha_n(f)$, $\beta_n(f)$ the angles of the vectors x_n, y_n, $f(x_n)$, $f(y_n)$. Then in the case (a) we have

$$|\alpha_n-\beta_n|_{2\pi} \ll |\alpha_n(f)-\beta_n(f)|_{2\pi}$$

and in case (b)

$$|\alpha_n(f)-\beta_n(f)|_{2\pi} \ll |\alpha_n-\beta_n|_{2\pi},$$

where $|z_1-z_2|_{2\pi}$ denotes the mod 2π distance of z_1 and z_2 and $a \ll b$ means $|\frac{a_n}{b_n}| \to 0$. We use the following

LEMMA. *Let* x_n, $y_n \in \mathbf{R}^2$, $|\frac{x_n}{y_n}| \to 1$. *Then*

$$|x_n-y_n| \leqq c\Big||x_n|-|y_n|\Big| \;\Leftrightarrow\; |\alpha_n-\beta_n| \leqq c\left|\ln\left|\frac{x_n}{y_n}\right|\right|.$$

Using this, prove first that $f \in \mathrm{Lip}1$. Let first $0 \leqq \alpha_n, \beta_n \leqq \pi$. Then by definition

$$\alpha_n(f)-\beta_n(f)=\frac{\beta_n-\alpha_n}{\pi}\left(\left\{(-1)^{\left[\frac{\ln x_n}{2\pi}\right]}\ln|x_n|\right\}_{2\pi}+\frac{\pi}{2}\cos\left(\frac{\ln|x_n|}{2}\right)\right)$$

$$\text{(i)} \qquad +\frac{\pi-\beta_n}{\pi}\left(\left\{(-1)^{\left[\frac{\ln|y_n|}{2\pi}\right]}\ln|y_n|\right\}_{2\pi}-\left\{(-1)^{\left[\frac{\ln|x_n|}{2\pi}\right]}\ln|x_n|\right\}_{2\pi}\right)$$

$$+\frac{\beta_n}{2}\left(\cos\left(\frac{\ln|y_n|}{2}\right)-\cos\left(\frac{\ln|x_n|}{2}\right)\right).$$

Consequently (using $\beta_n-\alpha_n \to 0$)

$$\Big|\alpha_n(f)-\beta_n(f)\Big|_{2\pi} \leqq 4|\beta_n-\alpha_n|_{2\pi}+10\left|\ln\left|\frac{x_n}{y_n}\right|\right|.$$

By $|\alpha_n-\beta_n|_{2\pi} \ll |\alpha_n(f)-\beta_n(f)|_{2\pi}$, this implies

$$|\alpha_n-\beta_n|_{2\pi} \leqq |\alpha_n(f)-\beta_n(f)|_{2\pi} \leqq 20\left|\ln\left|\frac{x_n}{y_n}\right|\right|, \quad (n \geqq n_0)$$

and then our Lemma implies

$$|f(x_n)-f(y_n)| \leqq c\Big||f(x_n)|-|f(y_n)|\Big| = c\Big||x_n|-|y_n|\Big| \leqq c|x_n-y_n|,$$

a contradiction. The case $-\pi \leqq \alpha_n, \beta_n \leqq 0$ goes along the same lines. If $0 \leqq \alpha_n \leqq \pi$ and $-\pi \leqq \beta_n \leqq 0$, then there are two cases:

(a_1) $\qquad \alpha_n \to 0+, \quad \beta_n \to 0+;$

(a_2) $\qquad \alpha_n \to \pi-, \quad \beta_n \to -\pi+.$

Here we have instead of (i)

$$\alpha_n(f) - \beta_n(f) = -\frac{\alpha_n + \beta_n}{\pi}\left(\left\{(-1)^{\left[\frac{\ln x_n}{2\pi}\right]} \ln|x_n|\right\}_{2\pi} + \frac{\pi}{2}\cos\frac{\ln|x_n|}{2}\right) -$$

(ii) $$-\frac{\pi + \beta_n}{\pi}\left(\left\{(-1)^{\left[\frac{\ln|y_n|}{2\pi}\right]} \ln|y_n|\right\}_{2\pi} - \left\{(-1)^{\left[\frac{\ln|x_n|}{2\pi}\right]} \ln|x_n|\right\}_{2\pi}\right)$$

$$-2\pi - \frac{\beta_n}{2}\left(\cos\left(\frac{\ln|y_n|}{2}\right) - \cos\left(\frac{\ln|x_n|}{2}\right)\right).$$

In case (a_1) we get $|\alpha_n(f) - \beta_n(f)|_{2\pi}$ by subtracting 2π and then

$$|\alpha_n(f) - \beta_n(f)| \leqq c|\alpha_n - \beta_n|,$$

contradiction. In case (a_2) we have $|\alpha_n - \beta_n| = |\alpha_n + \beta_n| \geqq |\beta_n + \pi|$, hence we get

$$|\alpha_n(f) - \beta_n(f)| \leqq c|\alpha_n - \beta_n|_{2\pi} + c\left|\ln\frac{|x_n|}{|y_n|}\right|,$$

which can be finished as above. Hence f is Lipschitzian, indeed.

Now consider (b). Let first $0 \leqq \alpha_n, \beta_n \leqq \pi$, then as we have remarked,

(*) $$\frac{\pi}{2} \leqq \left\{(-1)^{\left[\frac{\ln|x_n|}{2}\right]} \ln|x_n|\right\} + \frac{\pi}{2}\cos\frac{\ln|x_n|}{2} \leqq \frac{3\pi}{2}$$

(except for $x_n = e^{(4k+2)\pi}$).

Since $|\alpha_n(f) - \beta_n(f)|_{2\pi} \ll |\alpha_n - \beta_n|_{2\pi}$, hence we get from (1) that $|\alpha_n(f) - \beta_n(f)| = |\alpha_n(f) - \beta_n(f)|_{2\pi}$ and then $|\beta_n - \alpha_n| = |\beta_n - \alpha_n|_{2\pi} \leqq c\left|\ln\left|\frac{c_n}{y_n}\right|\right|$ so

$$|x_n - y_n| \leqq c\left||x_n - y_n|\right| = c\left||f(x_n)| - |f(y_n)|\right| \leqq c\,|f(x_n) - f(y_n)|,$$

contradiction. In case $-\pi \leqq \alpha_n, \beta_n \leqq 0$ we have

$$\alpha_n(f) - \beta_n(f) = \frac{\alpha_n - \beta_n}{\pi}\left(\left\{(-1)^{\left[\frac{\ln x_n}{2\pi}\right]} \ln|x_n|\right\}_{2\pi} - 2\pi + \frac{\pi}{2}\cos\frac{\ln|x_n|}{2}\right) -$$

(iii) $$-\frac{\pi + \beta_n}{\pi}\left((-1)^{\left[\frac{\ln|y_n|}{2\pi}\right]} \ln|y_n| - \left\{(-1)^{\left[\frac{\ln|x_n|}{2}\right]} \ln|x_n|\right\}_{2\pi}\right) -$$

$$-\frac{\beta_n}{2}\left(\cos\frac{\ln|y_n|}{2} - \cos\frac{\ln|x_n|}{2}\right).$$

Since (except for $|x_n| = e^{(4k+2)\pi}$)

$$-\frac{3\pi}{2} \leqq \left\{(-1)^{\left[\frac{\ln|x_n|}{2}\right]} \ln|x_n|\right\}_{2\pi} - 2\pi + \frac{\pi}{2}\cos\frac{\ln|x|}{2} \leqq \frac{\pi}{2}$$

hence we can finish this case as the case $0 \leqq \alpha_n, \beta_n \leqq \pi$.

Finally let $0 \leqq \alpha_n \leqq \pi$, $-\pi \leqq \beta_n \leqq 0$. The equality (ii) can be rewritten as

$$(**) \qquad \alpha_n(f) - \beta_n(f) = -\frac{\alpha_n}{\pi}t + \frac{\beta_n}{\pi}(2\pi - t) + O\left(\left|\ln\left|\frac{x_n}{y_n}\right|\right|\right) + 2\pi,$$

where

$$\frac{\pi}{2} \leqq \left\{(-1)^{\left[\frac{\ln|x_n|}{2}\right]} \ln|x_n|\right\}_{2\pi} + \frac{\pi}{2}\cos\frac{\ln|x_n|}{2} \leqq \frac{3\pi}{2}.$$

Since

$$0 \geqq -\frac{\alpha_n}{\pi}t + \frac{\beta_n}{\pi}(2\pi - t) \geqq -2\pi$$

and since $\alpha_n(f) - \beta_n(f) \to 0$, we have two possibilities

$$(b_1) \qquad \alpha_n \to 0+, \quad \beta_n \to 0-$$

$$(b_2) \qquad \alpha_n \to \pi-, \quad \beta_n \to -\pi+.$$

In case (b_1) we have again by $(**)$ and by $|\alpha_n(f) - \beta_n(f)|_{2\pi} \ll |\alpha_n - \beta_n|_{2\pi}$ that $|\alpha_n - \beta_n| \leqq c\left|\ln\left|\frac{x_n}{y_n}\right|\right|$ and $|\alpha_n(f) - \beta_n(f)|_{2\pi} \leqq c\left|\ln\left|\frac{x_n}{y_n}\right|\right|$ and then

$$|x_n - y_n| \leqq c\Big||x_n| - |y_n|\Big| = c\Big||f(x_n)| - |f(y_n)|\Big| \leqq c|f(x_n) - f(y_n)|.$$

Finally, in case (b_2) we have

$$|\alpha_n(f) - \beta_n(f)|_{2\pi} = \left|-\frac{\alpha_n}{\pi}t + \frac{\beta_n}{\pi}(2\pi - t) + 2\pi\right| + O\left(\left|\ln\left|\frac{x_n}{y_n}\right|\right|\right) =$$

$$= \left|\frac{\pi - \alpha_n}{\pi}t + \frac{\beta_n + \pi}{\pi}(2\pi - t)\right| + O\left(\left|\ln\left|\frac{x_n}{y_n}\right|\right|\right) \leqq$$

$$\leqq \left(|\alpha_n - \beta_n|_{2\pi} + \left|\ln\left|\frac{x_n}{y_n}\right|\right|\right)$$

and we are ready with the proof of $f^{-1} \in \mathrm{Lip}1$.

REMARK 1. Theorem (1) remains valid for large dimensions $N > 2$ (for $f : \mathbf{R}^N \to \mathbf{R}^N$), too. Contrarily to the case $N = 2$ this is simple to see: if we are given a curve $t \mapsto f(x + te)$ which meets infinitely often every direction for $t \to 0+$, we can move it at a distance $< \varepsilon$ such that this property fails. We do not discuss the details.

REMARK 2. In Theorem (2) we take Peano-type curves (i.e. curves whose central projection to the unit ball surface across every point infinitely often if $t \to 0+$). We rotate this curve to obtain a function $f : \mathbf{R}^N \to \mathbf{R}^N$ whose Lipschitz property will easily follow from the construction. The main difficulty will be the verification of the Lipschitz property of f^{-1}. The methods are similar to the above ones, hence we omit the details.

REMARK 3. Several other questions can be posed.
Problem 1. Construct $f : \mathbf{R}^2 \to \mathbf{R}^2$ such that $Df(x, e_0) = \mathbf{R}^2$.
Problem 2. Can it be for all $|e_0| = 1$?

REFERENCES

[1] JOÓ, I., On the divergence of eigenfunction expansions, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **32** (1989), 3–36.
[2] HORVÁTH, M., On multidimensional universal functions, *Studia Sci. Math. Hungar.* **22** (1987), 75–78. *MR* **88m**: 26013
[3] BOGMÉR, A. and SÖVEGJÁRTÓ, A., On universal functions, *Acta Math. Hungar.* **49** (1987), 237–239. *MR* **88a**: 26006
[4] BUCZOLICH, Z., On universal functions and series, *Acta Math. Hungar.* **49** (1987), 403–414. *MR* **88k**: 42011

(Received November 24, 1989)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

3414 STANFIELD DRIVE
PARMA, OH 44134
U.S.A

Current address:

WELDING ENGINEER TOOL SERVICES DEPARTMENT
BODY AND ASSEMBLY OPERATIONS
1700 OAKWOOD BOULEVARD
P.O.BOX 1586, ROOM E-1545
DEARBORN, MI 48121
FAX: (313) 845–6117
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 267–275

COMPOSITIONS OF AN INTEGER AND DISTRIBUTIONS OF RANK ORDER STATISTICS

JAGDISH SARAN

Abstract

This paper deals with the two-sample (equal sized) problem where $F_n(x)$ and $G_n(x)$ are the two empirical distribution functions and investigates the null joint and marginal distributions of certain rank order statistics, viz. β_n, the number of equalizations between $F_n(x)$ and $G_n(x)$; β_n^+, the number of equalizations between $F_n(x)$ and $G_n(x)$ from positive side; λ_n, the number of intersections between $F_n(x)$ and $G_n(x)$; and R_n, the number of runs in the ordered pooled sample by using combinatorial methods and generating functions, thus generalizing and extending the earlier work due to Csáki and Vincze [1], Jain [3], Kanwar Sen [4, 5] and Srivastava [10]. Also the interpretation of some of these results in terms of compositions of a positive integer n has been given.

1. Introduction

In [7], Narayana has considered a generalized occupancy problem which can be viewed as a problem in compositions of integers. Narayana and Fulton [9] considered the r-composition (or r-partition) of a positive integer n $(1 \leqq r \leqq n)$ and discussed its various properties. Also they discussed the relation of 'Domination' defined on the r-compositions of n, which is reflexive, transitive and antisymmetric. Thus it represents a 'Partial Order' defined on the r-compositions of n.

Narayana [8] discussed the same domination principle and the partial order defined on the compositions of an integer and gave some of its applications in probability theory. He gave a geometric representation of the r-compositions of n and proved that the number of r-compositions of n dominated by the r-compositions of n is given by

$$\frac{1}{n}\binom{n}{r}\binom{n}{r-1}. \tag{1}$$

In the terminology of 'lattice paths', this is equivalent to the number of lattice paths from $(0,0)$ to (n,n) starting with a horizontal step and never

1980 *Mathematics Subject Classification* (1985 Revision). Primary 62G30.

Key words and phrases. Two-sample problem, random walk, rank order statistics, sojourn, positive sojourn, intersection, run, partition of a positive integer, domination, strict domination.

Akadémiai Kiadó, Budapest

crossing the line $y = x$, each path having exactly r horizontal and r vertical components. Clearly, both horizontal and vertical components of each path represent an r-composition of n. Also, in terms of the usual 'Random walk model', this is equivalent to the number of random walk paths from $(0, 0)$ to $(2n, 0)$ starting with a positive step, having $2r$ runs and never crossing x-axis. Using the above mentioned approach of 'partial orders' defined on compositions of an integer, Mohanty and Narayana [6] gave two simple alternative solutions to 'ballot problems' [2, p. 69] in probability.

Srivastava [10] derived the joint distribution and the joint limiting distribution of the statistic based on the number of runs and that based on the number of intersections by using the method of composed paths. In this paper, we shall consider the two-sample problem and the random walk model as considered by Csáki and Vincze [1], Jain [3], Kanwar Sen [4, 5] and Srivastava [10] and propose to derive, under $H_0 : F(x) = G(x)$, the joint distributions of statistics based on the number of runs, returns, positive returns and intersections by using combinatorial methods and generating functions, thus generalizing and extending the earlier works in [1, 3, 4, 5, 10, 11]. Further we give the interpretation of some of these results in terms of the compositions of a positive integer n.

2. Notations

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ denote random samples drawn from populations with unknown continuous distribution functions $F(x)$ and $G(x)$, respectively. Let $F_n(x)$ and $G_n(x)$ be the corresponding empirical distribution functions. Let the $2n$ random variables be arranged in increasing order and denote the ordered combined sample by $Z_1 < Z_2 < \ldots < Z_{2n}$ and let $Z_0 = -\infty$. Now we introduce the random variables

$$\Theta_i = \begin{cases} +1 & \text{if } Z_i \text{ is one of the observations } X_1, X_2, \ldots, X_n \\ -1 & \text{if } Z_i \text{ is one of the observations } Y_1, Y_2, \ldots, Y_n \end{cases}$$

$i = 1, 2, \ldots, 2n$. This new sequence of n $(+1)$'s and n (-1)'s is called the sequence of rank order indicators. Under $H_0 : F(x) = G(x)$, there are $\binom{2n}{n}$ equally likely sequences of rank order indicators. A random variable which is a function of X's and Y's only through these rank order indicators is called the rank order statistic. Generally, rank order statistics are defined in terms of $F_n(x)$ and $G_n(x)$. Alternatively, such statistics can also be defined in terms of the partial sums S_i of the random walk $\{S_i\}$ generated by the sequence of independent random variables $\{\Theta_i\}$ defined below.

Let

$$S_i = \Theta_1 + \Theta_2 + \ldots + \Theta_i, \quad i = 1, 2, \ldots, 2n$$

with $S_0 = 0 = S_{2n}$. If the points (i, S_i) are plotted in a plane and each one of them is connected with the next one, we obtain the usual illustrative figure

of random walk path starting at the origin and returning after $2n$ steps to the origin. The array $(S_0, S_1, \ldots, S_{2n})$ is called the path of the particle for the two-sample problem of size n each.

The statistical problem in question is to ascertain whether or not two samples are from the same population (i.e., $F = G$), and thus it is important to derive probability distributions of various statistics when $H_0 : F(x) = G(x)$ is true. Some rank order statistics follow whose distributions, under H_0, will be derived in the sequel:

β_n, the number of returns to the origin.
$\beta_n =$ the number of points (i, S_i) such that $S_i = 0$.
β_n^+, the number of positive returns to the origin.
$\beta_n^+ =$ the number of points (i, S_i) such that $S_i = 0$ and $S_{i-1} = +1$.
λ_n, the number of intersections of the origin.
$\lambda_n =$ the number of points (i, S_i) such that $S_i = 0$ and $S_{i-1}S_{i+1} = -1$.
$=$ the number of changes of sign in $S_1, S_2, \ldots, S_{2n}$.
R_n, the number of runs in the sequence $\Theta_1, \Theta_2, \ldots, \Theta_{2n}$.
$R_n = 1+$ (the number of changes of sign in $\Theta_1, \Theta_2, \ldots, \Theta_{2n}$).

For simplicity of writing we introduce the following symbols:

F_{2n}: a path from $(0,0)$ to $(2n,0)$.
V-point: a point (i, S_i) of an F_{2n} path for which $S_i = 0$, we call it a return to the origin. The point $(0,0)$ is not regarded as a V-point.
$V^+(V^-)$: a V-point (i, S_i) such that $S_{i-1} = +1$ $(S_{i-1} = -1)$ and is called a positive (negative) return.
$W =$ wave: the segment of a path included between two consecutive V-points is called a wave. The path segment between the origin and the first V-point is also regarded as a wave.
$W^+(W^-)$: a wave (W) with $S_i > 0$ $(S_i < 0)$ at the intervening positions and is called a positive (negative) wave.
T-point: a point (i, S_i) of an F_{2n} path for which $S_i = 0$, $S_{i-1}S_{i+1} = -1$. This is called the intersection point of the x-axis.
$S =$ section: the segment of a path included between two consecutive T-points is called a section. We also treat as sections two other segments of the path, viz. the one from the origin to the first T-point and that from the last T-point to the last V-point.
$S^+(S^-)$: a section (S) with $S_i \geqq 0$ $(S_i \leqq 0)$ in-between.
$F_{2n}^{R,m}$: an F_{2n} path having R runs and m T-points.
$F_{2n}^{R+,m}(F_{2n}^{R-,m})$: an $F_{2n}^{R,m}$ path starting with a positive (negative) step.
$F_{2n}^{R+,m,p}(F_{2n}^{R-,m,p})$: an $F_{2n}^{R+,m}(F_{2n}^{R-,m})$ path having p V-points.
$F_{2n}^{R+,m,p,q}(F_{2n}^{R-,m,p,q})$: an $F_{2n}^{R+,m,p}(F_{2n}^{R-,m,p})$ path having q $(0 \leqq q \leqq p)$ V^+ points.

$F_{2n}^{r+,m,p}(F_{2n}^{r-,m,p})$: an F_{2n} path starting with a positive (negative) step and having m T-points and p V-points.

$F_{2n}^{r+,m,p,q}(F_{2n}^{r-,m,p,q})$: an $F_{2n}^{r+,m,p}(F_{2n}^{r-,m,p})$ path having q $(0 \leqq q \leqq p)V^+$ points.

$N(A)$: number of all possible paths of type A, e.g., $N(F_{2n}) = \binom{2n}{n}$.

3. Joint distributions based on β_n, β_n^+, λ_n and R_n

We note from [10; (3.6) for $j=0$] that

$$
\begin{aligned}
N\left(F_{2n}^{2r+,2m}\right) &= \frac{2m+1}{n}\binom{n}{r-m-1}\binom{n}{r+m} = \\
&= \binom{n-1}{r+m-1}\binom{n}{r-m-1} - \binom{n}{r+m}\binom{n-1}{r-m-2} = \\
&= \text{coeff. of } (yz)^n \text{ in} \\
&\left[\frac{y^{r+m}}{(1-y)^{r+m}}\frac{z^{r-m-1}}{(1-z)^{r-m}} - \frac{y^{r+m}}{(1-y)^{r+m+1}}\frac{z^{r-m-1}}{(1-z)^{r-m-1}}\right].
\end{aligned} \tag{2}
$$

An $F_{2n}^{2r+,2m}$ path, as shown in Fig. 1, consists of $(2m+1)S$, i.e., $(m+1)S^+$ and mS^- and has $2r$ runs. On changing the signs of Θ's of all those segments lying below the x-axis (as shown by dotted lines in Fig. 1) we get an $F_{2n}^{2(r+m)+,0}$ path having $(2m+1)S^+$. On using the reverse transformation we get back the original path and hence the transformation is one-to-one.

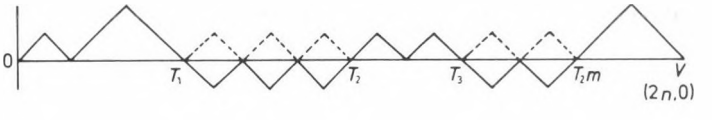


Fig. 1

Thus we can regard the expression on the right-hand side of (2) as the generating function of the number of paths from $(0,0)$ to $(2n,0)$ starting with a positive step, having $(2r+2m)$ runs, $(2m+1)S^+$ and never crossing x-axis.

Now on using the above result (2) we derive the following lemmas.

LEMMA 1.

$$
N(F_{2n}^{2r+,0,p}) = N(F_{2n}^{2r-,0,p}) = \frac{p}{r}\binom{n-p-1}{r-p}\binom{n-1}{r-1}. \tag{3}
$$

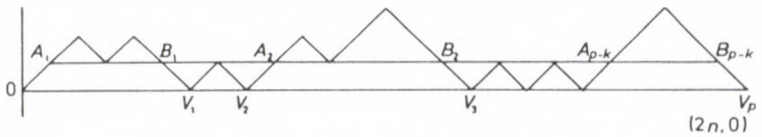


Fig. 2

PROOF. To derive (3), let $0A_1B_1V_1V_2 \dots A_{p-k}B_{p-k}V_p$ (Fig. 2) be an $F_{2n}^{2r+,0,p}$ path with $V_1, V_2, \dots, V_p$ as the p V-points. This path consists of p W^+. Let k $(k = 0, 1, \dots, p)$ out of p W^+ be of length two each, i.e., each having two runs as well. Let the remaining $(p-k)$ W^+ (each of length > 2) have $2r_1, 2r_2, \dots, 2r_{p-k}$ runs, respectively, such that $\sum_{j=1}^{p-k} r_j = r - k$. Draw a line $y = 1$ (see Fig. 2) and remove the portions of the path between $y = 0$ and $y = 1$. On joining the remaining segments of the path end-to-end, in order, we get an $F_{2n-2p}^{(2r-2k)+,0}$ path having $(p-k)$ S^+. Thus on using (2) we can write the generating function of $N(F_{2n}^{2r+,0,p})$ by replacing therein r by $\frac{1}{2}(2r-p-k+1)$ and m by $\frac{1}{2}(p-k-1)$, i.e.,

$$N(F_{2n}^{2r+,0,p}) = \text{coeff. of } (yz)^n \text{ in}$$

$$\sum_{k=0}^{p} \binom{p}{k} (yz)^p \left[\frac{y^{r-k}z^{r-p}}{(1-y)^{r-k}(1-z)^{r-p+1}} - \frac{y^{r-k}z^{r-p}}{(1-y)^{r-k+1}(1-z)^{r-p}} \right] = \tag{4}$$

$$= \text{coeff. of } (yz)^n \text{ in}$$

$$\left[\frac{y^r z^r}{(1-y)^r (1-z)^{r-p+1}} - \frac{y^r z^r}{(1-y)^{r+1}(1-z)^{r-p}} \right].$$

The factor $\binom{p}{k}$ is taken due to the reason that there can be any k W^+, each of length two, out of p W^+ and $(yz)^p$ is the generating function of the portion of the path of length $2p$ between the lines $y = 0$ and $y = 1$. Hence

$$N(F_{2n}^{2r+,0,p}) = \binom{-r}{n-r}\binom{-r+p-1}{n-r} - \binom{-r-1}{n-r}\binom{-r+p}{n-r} =$$

$$= \binom{n-1}{r-1}\binom{n-p}{r-p} - \binom{n}{r}\binom{n-p-1}{r-p-1},$$

leading to the first part of (3). The second part of (3) is also valid due to symmetry.

DEDUCTIONS. (i) Putting $p = 1$ in (3), we get

$$N(F_{2n}^{2r+,0,1}) = N(F_{2n}^{2r-,0,1}) = \frac{1}{r}\binom{n-2}{r-1}\binom{n-1}{r-1}. \tag{5}$$

(ii) Summing (3) over p from 1 to r and using the summation formula in Feller [2; II (12.16)], it verifies (1).

(iii) Summing (3) over r from p to n and using [2; II (12.9)], we get

$$N(F_{2n}^{+,0,p}) = N(F_{2n}^{-,0,p}) = \frac{p}{2n-p}\binom{2n-p}{n},$$

verifying a result in Feller [2, p. 90].

REMARK. Narayana [8] has defined that $(t_1, t_2, \ldots, t_r)$ is an r-composition of a positive integer n $(1 \leqq r \leqq n)$ if and only if

$$\sum_{i=1}^{r} t_i = n \text{ and } t_i \geqq 1, \quad i = 1, 2, \ldots, r.$$

Further, the r-composition $(t_1, t_2, \ldots, t_r)$ of n dominates another r-composition $(t_1', t_2', \ldots, t_r')$ of n if the following conditions hold:

$$\begin{aligned} t_1 &\geqq t_1' \\ t_1 + t_2 &\geqq t_1' + t_2' \\ t_1 + t_2 + t_3 &\geqq t_1' + t_2' + t_3' \\ &\vdots \\ t_1 + t_2 + \ldots + t_{r-1} &\geqq t_1' + t_2' + \ldots + t_{r-1}' \\ t_1 + t_2 + \ldots + t_r &= t_1' + t_2' + \ldots + t_r' = n. \end{aligned} \tag{6}$$

According to Narayana [8, p. 93], an r-composition of n dominated by another r-composition of n can be represented geometrically by a lattice path from $(0, 0)$ to (n, n) not rising above the line $y = x$ and having exactly r horizontal and r vertical components. Hence the expression in (1) is equivalent to the number of lattice paths from $(0, 0)$ to (n, n) never rising above the line $y = x$ and each having exactly r horizontal and r vertical components. Clearly both horizontal and vertical components of each path represent an r-composition of n.

Likewise we can interpret the right-hand side of (3) as the number of r-compositions of n dominated by r-compositions of n subject to the restriction that any $p-1$ relationships out of the first $r-1$ in (6) are equalities (so that the last relationship in (6) becomes the p-th equality) and the rest are strict inequalities. In other words, (3) is the number of lattice paths from $(0, 0)$ to (n, n) with r horizontal and r vertical components, never rising above the line $y = x$ and having exactly p contacts with $y = x$ (including the last one at (n, n)). In a similar manner we can interpret (5) as the number of r-compositions of n 'Strictly dominated' by r-compositions of n (strict domination means the $(r-1)$ inequalities in (6) are all strict inequalities).

In other words, (5) is the number of lattice paths from $(0,0)$ to (n,n) lying entirely below the line $y=x$, never touching it in-between except at the end points, each path having exactly r horizontal and r vertical components.

LEMMA 2.

$$(7)\quad N(F_{2n}^{2r+,2m,p,q}) = \frac{p}{r+m}\binom{q-1}{m}\binom{p-q-1}{m-1}\binom{n-p-1}{r+m-p}\binom{n-1}{r+m-1}$$

$$(8)\quad N(F_{2n}^{2r-,2m,p,q}) = \frac{p}{r+m}\binom{q-1}{m-1}\binom{p-q-1}{m}\binom{n-p-1}{r+m-p}\binom{n-1}{r+m-1}$$

and

$$(9)\quad \begin{aligned} N(F_{2n}^{(2r+1)+,2m+1,p,q}) &= N(F_{2n}^{(2r+1)-,2m+1,p,q}) = \\ &= \frac{p}{r+m+1}\binom{q-1}{m}\binom{p-q-1}{m}\binom{n-p-1}{r+m-p+1}\binom{n-1}{r+m}. \end{aligned}$$

PROOF. An $F_{2n}^{2r+,2m,p,q}$ path consists of $(m+1)$ S^+ (comprising q W^+) and m S^- (comprising $(p-q)$ W^-). On changing the signs of all those segments lying below the x-axis, we get an $F_{2n}^{2(r+m)+,0,p}$ path. Since $(m+1)$ S^+ and m S^- can be constructed out of q ordered W^+ and $(p-q)$ ordered W^- in $\binom{q-1}{m}\binom{p-q-1}{m-1}$ ways, we get (7) by using (3) where r is replaced by $r+m$. Others are similarly established.

DEDUCTIONS. (i) For $m=0$, (7) and (8) reduce to

$$N(F_{2n}^{2r+,0,p}) = N(F_{2n}^{2r-,0,p}) = \frac{p}{r}\binom{n-p-1}{r-p}\binom{n-1}{r-1},$$

since in this case q will be equal to p and 0 in (7) and (8), respectively, thus verifying result (3).

(ii) Summing (7) and (8) each over $p-m \leqq r \leqq n-m$ and (9) over $p-m-1 \leqq r \leqq n-m-1$ and using [2; II (12.9)], we get, respectively

$$N(F_{2n}^{+,2m,p,q}) = \frac{p}{2n-p}\binom{2n-p}{n}\binom{q-1}{m}\binom{p-q-1}{m-1},$$

$$N(F_{2n}^{-,2m,p,q}) = \frac{p}{2n-p}\binom{2n-p}{n}\binom{q-1}{m-1}\binom{p-q-1}{m},$$

and

$$N(F_{2n}^{+,2m+1,p,q}) = N(F_{2n}^{-,2m+1,p,q}) = \frac{p}{2n-p}\binom{2n-p}{n}\binom{q-1}{m}\binom{p-q-1}{m},$$

verifying known results in [3].

(iii) Summing (7) over $m+1 \leqq q \leqq p-m$, (8) over $m \leqq q \leqq p-m-1$ and (9) over $m+1 \leqq q \leqq p-m-1$ and using [2; II (12.16)], we get, respectively

$$(10) \quad N(F_{2n}^{2r+,2m,p}) = N(F_{2n}^{2r-,2m,p}) = \frac{p}{r+m}\binom{p-1}{2m}\binom{n-p-1}{r+m-p}\binom{n-1}{r+m-1}$$

and

$$(11) \quad N(F_{2n}^{(2r+1)+,2m+1,p}) = N(F_{2n}^{(2r+1)-,2m+1,p}) = \\ = \frac{p}{r+m+1}\binom{p-1}{2m+1}\binom{n-p-1}{r+m-p+1}\binom{n-1}{r+m}.$$

(iv) Setting $m=0$ in (10), it verifies (3).

(v) Summing (10) over $2m+1 \leqq p \leqq r+m$ and (11) over $2m+2 \leqq p \leqq r+m+1$ and using [2; II (12.16)], we get, respectively

$$N(F_{2n}^{2r+,2m}) = N(F_{2n}^{2r-,2m}) = \frac{2m+1}{n}\binom{n}{r-m-1}\binom{n}{r+m}$$

and

$$N(F_{2n}^{(2r+1)+,2m+1}) = N(F_{2n}^{(2r+1)-,2m+1}) = \frac{2m+2}{n}\binom{n}{r-m-1}\binom{n}{r+m+1},$$

verifying [10; (3.6) for $j=0$], [10; (3.10)], [10; (3.8)] and [10; (3.12) for $j=0$], respectively.

(vi) Summing (10) over $p-m \leqq r \leqq n-m$ and (11) over $p-m-1 \leqq r \leqq n-m-1$ and using [2; II (12.9)], we get, respectively

$$N(F_{2n}^{*+,2m,p}) = N(F_{2n}^{*-,2m,p}) = \frac{p}{2n-p}\binom{p-1}{2m}\binom{2n-p}{n}$$

and

$$N(F_{2n}^{*+,2m+1,p}) = N(F_{2n}^{*-,2m+1,p}) = \frac{p}{2n-p}\binom{p-1}{2m+1}\binom{2n-p}{n},$$

verifying known results in [3].

The foregoing lemmas lead immediately to the following joint distributions.

THEOREM.

$$P[\lambda_n = 0, \beta_n = p, R_n = 2r] = \frac{2p}{n}\binom{n-p-1}{r-p}\binom{n}{r}\Big/\binom{2n}{n},$$

$$P[\lambda_n = 2m, \beta_n = p, R_n = 2r] = \frac{2p}{n}\binom{p-1}{2m}\binom{n-p-1}{r+m-p}\binom{n}{r+m}\Big/\binom{2n}{n},$$

$$P[\lambda_n = 2m+1, \beta_n = p, R_n = 2r+1] =$$

$$= \frac{2p}{n}\binom{p-1}{2m+1}\binom{n-p-1}{r+m-p+1}\binom{n}{r+m+1}\Big/\binom{2n}{n},$$

$$P\left[\lambda_n = 2m, \beta_n = p, \beta_n^+ = q, R_n = 2r\right] =$$

$$= \frac{p(p-2m)}{nq}\binom{q}{m}\binom{p-q-1}{m-1}\binom{n-p-1}{r+m-p}\binom{n}{r+m}\Big/\binom{2n}{n}$$

and

$$P\left[\lambda_n = 2m+1, \beta_n = p, \beta_n^+ = q, R_n = 2r+1\right]$$

$$= \frac{2p}{n}\binom{q-1}{m}\binom{p-q-1}{m}\binom{n-p-1}{r+m-p+1}\binom{n}{r+m+1}\Big/\binom{2n}{n}.$$

REFERENCES

[1] Csáki, E. and Vincze, I., On some problems connected with the Galton-test, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **6** (1961), 97–109. *MR* **26** # 3138

[2] Feller, W., *An introduction to probability theory and its applications*, Vol. I, 3rd ed., John Wiley, New York, 1968. *MR* **37** # 3604

[3] Jain, G. C., Joint distributions of intersections, (±) waves and (±) steps I, *Proc. Nat. Inst. Sci. India Part A* **32** (1966), 460–471. *MR* **41** # 6317

[4] Sen, K., On some combinatorial relations concerning the symmetric random walk, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **9** (1964), 335–357. *MR* **33** # 6715

[5] Sen, K., Paths of an odd number of steps with final position unspecified, *J. Indian Statist. Assoc.* **7** (1969), 107–135. *MR* **52** # 1896

[6] Mohanty, S. G. and Narayana, T. V., Some properties of compositions and their application to probability and statistics I, *Biometr. Z.* **3** (1961), 252–258. *Zbl* **101**, 359

[7] Narayana, T. V., A combinatorial problem and its application to probability theory I, *Indian Soc. Agric. Statist.* **7** (1955), 169–178. *MR* **19**–523

[8] Narayana, T. V., A partial order and its applications to probabiliy theory, *Sankhya* **21** (1959), 91–98. *MR* **21** # 5230

[9] Narayana, T. V. and Fulton, G. E., A note on the compositions of an integer, *Canad. Math. Bull.* **1** (1958), 169–173. *MR* **21** # 2634

[10] Srivastava, S., Joint distributions based on runs and on the number of intersections, *Studia Sci. Math. Hungar.* **8** (1973), 211–224. *MR* **50** # 11464

[11] Saran, J. and Sen, K., On the joint distribution of some rank order statistics, *Statistica* **40** (1980), 235–240. *MR* **81i**:62079

(Received December 15, 1989)

DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICAL SCIENCES
UNIVERSITY OF DELHI
IND–110007 DELHI
INDIA

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 277–283

COVERING A DISK WITH SMALLER DISKS

S. KROTOSZYŃSKI

Let D be a disk in the Euclidean plane. By $D(k)$ we denote the smallest positive ratio of k homothetical copies of D whose union covers D. The following values of $D(k)$ are known: $D(1) = D(2) = 1$, $D(3) = \frac{\sqrt{3}}{2}$, $D(4) = \frac{\sqrt{2}}{2}$, $D(5) = 0.609\ldots$, $D(6) = 0.556\ldots$, and $D(7) = \frac{1}{2}$ (see [1–4]).

In this paper an index is understood as corresponding number between 1 and $k-1$ modulo $k-1$. The boundary and the interior of a plane convex body A are denoted by $\operatorname{bd} A$ and $\operatorname{int} A$, respectively. The segment with endpoints a and b is denoted by ab, and its length is denoted by $|ab|$. The distance of sets A and B is denoted by $d(A, B)$.

THEOREM. *If* $k \in \{8, 9, 10, 11\}$, *then* $D(k) = \left(1 + 2\cos\frac{2\pi}{k-1}\right)^{-1}$.

For the proof we need the following lemmas.

LEMMA 1. *Let k be a positive integer greater than 7. If $\beta_i > 0$ and $\beta_{i-1} + \beta_i + \beta_{i+1} < \pi$ for $i = 1, \ldots, k-1$ and if $\beta_1 + \ldots + \beta_{k-1} = 2\pi$, then*

$$\sum_{i=1}^{k-1} [\sin\beta_{i-1} + \sin(\beta_i + \beta_{i+1})] \leqq$$

$$\leqq \sum_{i=1}^{k-1} \left[\sin\frac{\beta_{i-1} + \beta_i + \beta_{i+1}}{3} + \sin\frac{2(\beta_{i-1} + \beta_i + \beta_{i+1})}{3}\right].$$

The equality holds if and only if $\beta_1 = \ldots = \beta_{k-1} = 2\pi/(k-1)$.

LEMMA 2. *Let k be a positive integer greater than 7. If $0 < \gamma_i < \pi$ for $i = 1, \ldots, k-1$ and if $\gamma_1 + \ldots + \gamma_{k-1} = 6\pi$, then*

$$\sum_{i=1}^{k-1} \left[\frac{1}{2}\sin\gamma_i + \sin\frac{\gamma_i}{3} + \sin\frac{2\gamma_i}{3}\right] \leqq$$

$$\leqq \sum_{i=1}^{k-1} \sqrt{\sin^2\frac{1}{2}\gamma_i \left(1 + 2\cos\frac{2\pi}{k-1}\right)^2 - \sin^4\frac{1}{2}\gamma_i}$$

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 54A45.
Key words and phrases. Covering, disk, convex polygon.

Akadémiai Kiadó, Budapest

with equality if and only if $\gamma_1 = \ldots = \gamma_{k-1} = 2\pi/(k-1)$.

PROOF OF LEMMA 1. We present the proof only for $k = 8$ because for $k \in \{9, 10, 11\}$ the procedure is analogical.

Let us consider the function

$$F(x_1, \ldots, x_7) = \sum_{i=1}^{k-1} \left[\sin \frac{x_{i-1}+x_i+x_{i+1}}{3} + \sin \frac{2(x_{i-1}+x_i+x_{i+1})}{3}\right] - \tag{1}$$
$$- \sum_{i=1}^{k-1} [\sin x_{i-1} + \sin(x_i + x_{i+1})],$$

where $x_i > 0$ and $x_{i-1} + x_i + x_{i+1} < \pi$ for $i = 1, \ldots, 7$ and $x_1 + \ldots + x_{k-1} = 2\pi$.

Suppose that $x_i \leqq x_1$ for $i = 2, \ldots, 7$. We show that

$$\text{if} \quad \frac{\partial F}{\partial x_1} = 0, \quad \text{then} \quad x_1 = x_2 = x_3 = x_6 = x_7. \tag{2}$$

We have

$$\frac{\partial F}{\partial x_1} = \frac{1}{3}\left[\cos \frac{x_1+x_6+x_7}{3} + \cos \frac{x_1+x_2+x_7}{3} + \cos \frac{x_1+x_2+x_3}{3}\right] +$$
$$+ \frac{2}{3}\left[\cos \frac{2(x_1+x_6+x_7)}{3} + \cos \frac{2(x_1+x_2+x_7)}{3} + \cos \frac{2(x_1+x_2+x_3)}{3}\right] - \tag{3}$$
$$- [\cos x_1 + \cos(x_1 + x_7) + \cos(x_1 + x_2)].$$

The function $\cos\varphi$ is decreasing for φ belonging to the closed interval $[0, \pi]$. Hence and from $x_3 \leqq x_1$ and from $x_6 \leqq x_1$ we obtain

$$\frac{\partial F}{\partial x_1} \geqq \frac{1}{3}\left[\cos \frac{2x_1 + x_7}{3} + \cos \frac{2x_1 + x_2}{3} + \cos \frac{x_1 + x_2 + x_7}{3}\right] +$$
$$+ \frac{2}{3}\left[\cos \frac{2(2x_1 + x_7)}{3} + \cos \frac{2(2x_1 + x_2)}{3} + \cos \frac{2(x_1 + x_2 + x_7)}{3}\right] - \tag{4}$$
$$- [\cos x_1 + \cos(x_1 + x_2) + \cos(x_1 + x_7)].$$

We have the following inequalities:

$$\frac{1}{3}\cos \frac{2x_1 + x_2}{3} + \frac{2}{3}\cos \frac{2(2x_1 + x_2)}{3} \geqq \frac{1}{3}\cos x_1 + \frac{2}{3}\cos(x_1 + x_2), \tag{5}$$

$$\frac{1}{3}\cos \frac{2x_1 + x_7}{3} + \frac{2}{3}\cos \frac{2(2x_1 + x_7)}{3} \geqq \frac{1}{3}\cos x_1 + \frac{2}{3}\cos(x_1 + x_7), \tag{6}$$

$$\frac{1}{3}\cos \frac{x_1+x_2+x_7}{3} + \frac{2}{3}\cos \frac{2(x_1+x_2+x_7)}{3} \geqq \frac{1}{3}\cos x_1 + \frac{1}{3}\cos(x_1+x_2) + \tag{7}$$
$$+ \frac{1}{3}\cos(x_1 + x_7).$$

Let us show (5). For $0 \leqq x \leqq \frac{1}{3}x_1$ and $0 \leqq y \leqq x_1$ the function $H(x) = \frac{1}{3}\cos(x+y) + \frac{2}{3}\cos(y-x)$ attains the maximum for $x = 0$. Thus

$$\frac{1}{3}\cos y + \frac{2}{3}\cos 2y \geqq \frac{1}{3}\cos(x+y) + \frac{2}{3}\cos(y-x) \tag{8}$$

for every $x \in [0, \frac{1}{3}x_1]$ and every $y \in [0, x_1]$. Putting $x = \frac{1}{3}(x_1 - x_2)$ and $y = \frac{1}{3}(2x_1 + x_2)$ in (8) we obtain (5). Analogously we show (6) and (7). Adding inequalities (5), (6) and (7) we get $\frac{\partial F}{\partial x_1} \geqq 0$ with the equality only for $x_1 = x_2 = x_3 = x_6 = x_7$. This implies

$$\frac{\partial F}{\partial x_i} = 0 \quad \text{for} \quad i = 1, \ldots, 7 \quad \text{if and only if} \quad x_1 = x_2 = \ldots = x_7. \tag{9}$$

It is easy to check that the function F defined in (1) has the minimum equal to 0 for $x_1 = \ldots = x_7 = 2\pi/7$. Thus $F(x_1, \ldots, x_7) \geqq 0$ for $x_1, \ldots, x_7$ fulfilling the assumption of Lemma 1.

We omit an analogous proof of Lemma 2.

PROOF OF THE THEOREM. Let $k \in \{8, 9, 10, 11\}$. Consider the disk D of radius 1. Denote by s the centre of D. Let $D_1, \ldots, D_k$ be disks of radius $r_k = (1 + 2\cos\frac{2\pi}{k-1})^{-1}$. The symbol s_i means the centre of D_i for $i = 1, \ldots, k$. Let $\mathcal{D}_k = \{D_1, \ldots, D_k\}$.

For the proof of our Theorem we will show the following five statements:

(a) There are positions of $D_1, \ldots, D_k$ such that $D_1 \cup \ldots \cup D_k \supset D$.

(b) If $D_1 \cup \ldots \cup D_k \supset D$, then there exists a unique $i_0 \in \{1, \ldots, k\}$ such that $D_{i_0} \subset \operatorname{int} D$.

(c) If $D_1 \cup \ldots \cup D_k \supset D$, then there are translations $T_1, \ldots, T_k$ such that the intersection of each pair of the disks $T_i(D_i)$, for $i = 1, \ldots, k$, has at most one point in bd D.

(d) If $D_1 \cup \ldots \cup D_k \supset D$ and if each three disks of $\mathcal{D}_k$ have at most one point in common, then $D_1, \ldots, D_k$ coincide with disks described in (a).

(e) If $D_1 \cup \ldots \cup D_k \supset D$ and if the intersection of some three disks of $\mathcal{D}_k$ consists of more than one point, then there exist a number $r'_k < r_k$ and disks $D'_1, \ldots, D'_k$ of radius r'_k such that $D'_1 \cup \ldots \cup D'_k \supset D$.

PROOF OF (a). If $s_1 = s$ and if $s_2, \ldots, s_k$ are the vertices of the regular $(k-1)$-gon with the centre s and such that every three disks of $\mathcal{D}_k$ have at most one point in common, then $D_1 \cup \ldots \cup D_k \supset D$ (see Fig. 1).

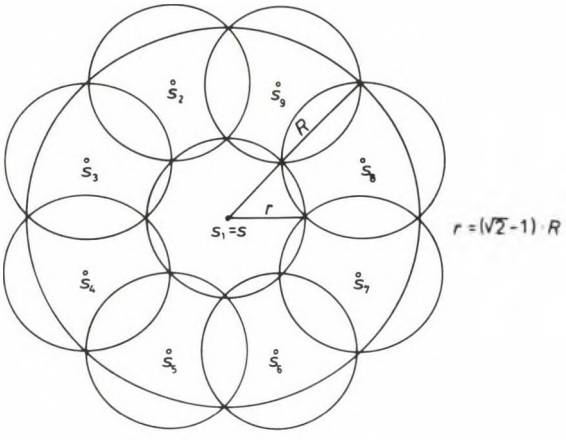


Fig. 1

Of course, the inclusion does not hold if we lessen the radius of $D_1, \ldots, D_k$.

PROOF OF (b). Let us assume that some m_k disks of $\mathcal{D}_k$ are subsets of int D. Then the union of the $k - m_k$ remaining disks covers bd D.

Observe that each of these disks is able to cover a part of bd D with the central angle not greater than $2\arcsin(1 + 2\cos\frac{2\pi}{k-1})^{-1}$. Hence

$$(k - m_k)\arcsin\left(1 + 2\cos\frac{2\pi}{k-1}\right)^{-1} \geqq \pi.$$

From this inequality it follows that $m_8 < 1.22$, $m_9 < 1.97$, $m_{10} < 2.27$ and $m_{11} < 2.99$. Since m_8 and m_9 are smaller than 2, we see that (b) holds true for $k = 8$ and $k = 9$.

Now, we consider the case when $k = 10$ (for $k = 11$ we proceed analogously). Let W be a regular octagon with the centre in s and with radius $r_w = 5.1$ of the circumscribed circle. It is not hard to see that if some two disks of $\mathcal{D}_{10}$ are subsets of int D, then they cover at most four sides of W. Since the union of the remaining disks of $\mathcal{D}_{10}$ is able to cover only three sides of W, it is impossible to cover the whole octagon W by the union of disks of $\mathcal{D}_{10}$.

Thanks to the above considerations we may assume that $D_k \subset \text{int } d$.

PROOF OF (c). Let $D_k \subset \text{int } D$ and let the points $s_1, \ldots, s_{k-1}$ be consecutive vertices of a $(k-1)$-gon, according to the orientation of the plane. The common point of D_i, bd D_{i+1} and bd D is denoted by a_i for $i = 1, \ldots, k-1$. By a_i' we denote the common point of bd D_i, D_{i+1} and bd D for $i = 1, \ldots, k-1$.

From $D_1 \cup \ldots \cup D_k \supset D$ it easily results that $s_1, \ldots, s_k$ are in int D. Really, if $s_i \notin \text{int } D$ for some $i \in \{1, \ldots, k-1\}$, then let $S(D_i)$ be the symmetric image of D_i with respect to the straight line through a_{i-1} and a_i'. Obviously,

the centre of $S(D_i)$ is in int D and the sum of $S(D_i)$ and the sets D_j for $j \in \{1, \ldots, k-1\}$ and $j \neq i$ covers D.

Moreover, for each $i = 1, \ldots, k-1$ we take the translations T_i of D_i such that bd $T_i(D_i)$ contains the points a_i and a_{i-1} and that the centre of $T_i(D_i)$ is in int D. Of course, $T_1(D_1) \cup \ldots \cup T_{k-1}(D_{k-1}) \cup D_k \supset D$ and $T_i(a_i') = a_i$ for $i = 1, \ldots, k-1$.

Proof of (d). We denote by b_i the point of intersection of bd D_i, bd D_{i+1} and bd D_k for $i = 1, \ldots, k-1$ and we denote by b_k the point of intersection of bd D_1, bd D_{k-1} and bd D_k.

Moreover, let $\alpha_i = \sphericalangle a_i s a_{i+1}$ and $\beta_i = \sphericalangle b_i s_k b_{i+1}$ for $i = 1, \ldots, k-1$. Hence $\beta_i + \beta_{i+1} = \sphericalangle s_{i+1} b_{i+1} s_{i+2}$ and $\beta_{i-1} + \beta_i + \beta_{i+1} = \sphericalangle a_i s_{i+1} a_{i+1}$ for $i = 1, \ldots, k-1$.

Let us consider the following figures: the rhombi with vertices s_1, b_1, s_{i+1}, b_{i+1} and b_i, s_i, a_i, b_{i+1} for $i = 1, \ldots, k-1$, the triangles with vertices a_i, s_{i+1}, a_{i+1} for $i = 1, \ldots, k-1$, the segments of the disk D with chords $a_i a_{i+1}$ for $i = 1, \ldots, k-1$. Since $D_1 \cup \ldots \cup D_k \supset D$, the sum of areas of these figures is equal to the area of D. Thus

$$\sum_{i=1}^{k-1} \left[\frac{\sin \beta_{i-1} + \sin(\beta_i + \beta_{i+1}) + \frac{1}{2} \sin(\beta_{i-1} + \beta_i + \beta_{i+1})}{\left(1 + 2 \cos \frac{2\pi}{k-1}\right)^2} + \frac{\alpha_i - \sin \alpha_i}{2} \right] = \pi.$$

Moreover, $\alpha_1 + \ldots + \alpha_{k-1} = 2\pi$ and

$$\sin \alpha_i = 2 \sqrt{r_k^2 \sin^2 \frac{1}{2}(\beta_{i-1} + \beta_i + \beta_{i+1}) - r_k^4 \sin^4 \frac{1}{2}(\beta_{i-1} + \beta_i + \beta_{i+1})}$$

for $i = 1, \ldots, k-1$. Hence we obtain the equality

$$\sum_{i=1}^{k-1} [\sin \beta_{i-1} + \sin(\beta_i + \beta_{i+1}) + \frac{1}{2} \sin(\beta_{i-1} + \beta_i + \beta_{i+1})] =$$

(10)

$$= \sum_{i=1}^{k-1} \sqrt{\left(1 + 2 \cos \frac{2\pi}{k-1}\right)^2 \sin^2 \frac{1}{2}(\beta_{i-1} + \beta_i + \beta_{i+1}) - \sin^4 \frac{1}{2}(\beta_{i-1} + \beta_i + \beta_{i+1})}.$$

From Lemmas 1 and 2 it follows that (10) holds if and only if $\alpha_1 = \ldots = \alpha_{k-1} = \beta_1 = \ldots = \beta_{k-1} = 2\pi/(k-1)$. Hence our covering coincides with the covering described in part (a).

Proof of (e). We consider the case $k = 9$. (For $k = 8, 10, 11$ we proceed analogously.)

Let the sets $D_1 \cap D_2 \cap D_9$ and $D_1 \cap D_8 \cap D_9$ have in common more than one point. Denote by x_1 the centre of gravity of $D_1 \cap D_2 \cap D_9$ and denote by x_2 the centre of gravity of $D_1 \cap D_8 \cap D_9$.

Let $t_1 \in \operatorname{int} D_1$ be such point of the ray from s through s_1 that $r_9 = \max\{|t_1x_1|, |t_1x_2|\}$. We provide the disk K_1 with the centre t_1 and with radius r_9. We denote by c_1 the common point of D_2, bd K_1 and bd D and we denote by c_8 the common point of D_8, bd K_1 and bd D. The point $t_2 \in D$ is the centre of the disk K_2 of radius r_9 such that $a_2, c_1 \in \operatorname{bd} K_2$ and the point $t_8 \in D$ is the centre of the disk K_8 of radius r_9 such that $a_7, c_8 \in \operatorname{bd} K_8$.

We denote by d_1 the point of intersection of D_9, bd K_1 and bd K_2 and we denote by d_2 the point of intersection of D_9, bd K_2 and bd D_3. Analogously, we denote by d_7 the point of intersection of D_9, bd D_7 and bd K_8 and we denote by d_8 the point of intersection of D_9, bd K_1 and bd D_8. Of course d_1, d_2, d_7 and d_8 are in int D_9.

We provide two tangents to the disk D_9 through the points b_3 and b_6. The point of intersection of these lines is denoted by p. Let us consider the ray from p through y, where $y \in b_3b_6$. For $i = 1, 2, 7, 8$ we denote by y_i the point of bd D_9 such that the vectors $\overrightarrow{d_iy_i}$ and $\overrightarrow{py}$ are parallel and that they have the same sense.

Let $\mathcal{H}_y = \min\{|d_1y_1|, |d_2y_2|, |d_7y_7|, |d_8y_8|\}$ and we define the number

$$\lambda = \max\{\mathcal{H}_y : y \in b_3b_6 \text{ and } b_3 \neq y \neq b_6\}.$$

It is easy to see that there exist $u \in b_3b_6$ and $j \in \{1, 2, 7, 8\}$ such that $\lambda = |b_ju_j|$. The point $w_9 \in \operatorname{int} D_9$ satisfying the condition $\overrightarrow{s_9w_9} = -\frac{1}{2}\overrightarrow{b_ju_j}$ is the centre of the disk L_9 of radius r_9. It is obvious that the points d_i for $i = 1, 2, 7, 8$ and the points b_i for $i = 3, 4, 5, 6$ are in int L_9.

Let $D_i = K_i$ for $i = 3, \ldots, 8$. We denote by v_i for $i = 1, \ldots, 8$ the centre of gravity of $K_i \cap K_{i+1} \cap L_9$. The point $w_i \in \operatorname{int} D_i$ for $i = 1, \ldots, 8$ is the centre of the disk L_i with radius r_9 such that w_i lies on the ray from s through s_i and that $r_9 = \max\{|w_iv_{i-1}|, |w_iv_i|\}$. Moreover, it is easy to check that the following sets: $X_i = L_{i-1} \cap L_i \cap L_9$ and $Y_i = L_{i-1} \cap L_i \cap \operatorname{bd} D$ for $i = 1, \ldots, 8$ (where $L_0 = L_8$) have more than one point.

We define the numbers

$$\delta_i = \min_{m \in X_i} \{\max(|w_{i-1}m|, |w_im|, |w_9m|)\}$$

and

$$\varepsilon_i = \min_{n \in Y_i} \{\max(|w_{i-1}n|, |w_in|)\}$$

for $i = 1, \ldots, 8$ (where $w_0 = w_8$). From the above considerations we conclude that $\delta_i < r_9$ and $\varepsilon_i < r_9$ for $i = 1, \ldots, 8$. Hence we obtain that the number

$$r = \max\{\delta_1, \ldots, \delta_8, \varepsilon_1, \ldots, \varepsilon_8\} \quad \text{is smaller than } r_9.$$

The disks $D_1', D_2', \ldots, D_9'$ with centres $w_1, \ldots, w_9$ and with radius r fulfil the inclusion $D_1' \cup \ldots \cup D_9' \supset D$. This ends the proof of (e).

From statements (a)–(e) it follows that r_k, where $k \in \{8, 9, 10, 11\}$, is the smallest number such that the disks $D_1, \ldots, D_k$ with radius r_k cover D of radius 1. Hence we have $D(k) = r_k$ for $k = 8, 9, 10, 11$. The proof is complete.

Finally, let us point out that

$$D(12) < \left(1 + 2\cos\frac{2\pi}{11}\right)^{-1} = 0.372\ldots.$$

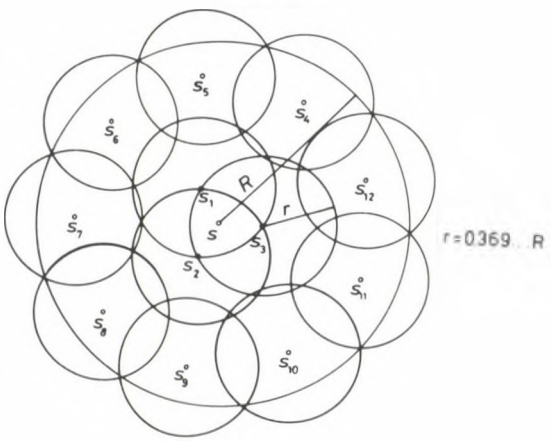


Fig. 2

Figure 2 shows disk D with the centre s and with the radius 1 covered by disks $D_1, \ldots, D_{12}$ of radius r. The centres of disks D_1, D_2, D_3 are the vertices of the regular triangle T with the midst in s. The length of the side of T is equal to r. The centres of disks $D_4, \ldots, D_{12}$ are the vertices of the regular nonagon with midst in s such that the following sets have one point in common: $D_1 \cap D_2 \cap D_7$, $D_2 \cap D_3 \cap D_{10}$, $D_1 \cap D_3 \cap D_4$.

It is easy to check that $r < (1 + 2\cos\frac{2\pi}{11})^{-1}$. Hence $D(12) < r < (1 + 2\cos\frac{2\pi}{11})^{-1}$.

REFERENCES

[1] Bezdek, K., Über einige Kreisüberdeckungen, *Beiträge Algebra Geom.* **14** (1983), 7–13. *MR* **85a**:52012

[2] Lassak, M., Covering plane convex bodies with smaller homothetical copies, *Intuitive geometry* (Siófok, 1985), Colloq. Math. Soc. J. Bolyai, Vol. 48, North-Holland, Amsterdam–New York, 1987, 331–337. *MR* **88i**:52023

[3] Molnár, J., Über eine elementargeometrische Extremalaufgabe, *Mat. Fiz. Lapok* **49** (1942), 249–253. *MR* **8**–218

[4] Neville, H., On the solution of numerical functional equations, *Proc. London Math. Soc.* (2) **14** (1915), 308–326. *Jb. Fortschritte Math.* **45**, 1230

(Received January 4, 1990)

INSTYTUT MATEMATYKI I FIZYKI ATR
AL. PROF. S. KALISKIEGO 7
PL–85-790 BYDGOSZCZ
POLAND

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 285–288

RINGS WITH LOCAL UNITS AND DESCENDING CHAIN CONDITION

PHAM NGOC ANH

Answering a question of Szász ([4], Problem 8 and Problem 9) in [1] we characterized primary rings with descending chain condition on finitely generated left ideals. The aim of the present paper is to describe direct sums of these rings and to discuss some interesting special cases. Our results are motivated by the investigations of Numakura [2], [3].

Recall (see e.g. [1]) that a ring R is said to have *local units* if for any finite subset A of R there is an idempotent $e \in R$ which acts as an identity on A on both sides. A ring R is said to be *primary* if it has local units and its factor by the radical is a simple ring. For any ideal I of an arbitrary ring R we define by transfinite induction the ideals

$$\underset{\mu+1}{I} = \underset{\mu}{I} \cdot \underset{\mu}{I}, \qquad I_{\mu+1} = I_\mu \cdot I,$$

and

$$\underset{\mu}{I} = \bigcap_{\alpha<\mu} \underset{\alpha}{I}, \qquad I_\mu = \bigcap_{\alpha<\mu} I_\alpha$$

if μ is a limit ordinal. Clearly, there is an ordinal α with $I_\alpha = I_\beta$ and $\underset{\alpha}{I} = \underset{\beta}{I}$ for all $\beta > \alpha$. These I_α and $\underset{\alpha}{I}$ will be denoted by I_* and $\underset{*}{I}$, respectively. Moreover, they satisfy $I_* \cdot I = I_*$ and $\underset{*}{I}^2 = \underset{*}{I}$. An ideal I is called *transfinitely nilpotent* if $I_* = 0$. It is well known that the radical of a ring with descending chain condition on finitely generated left ideals is transfinitely nilpotent. First we prove:

THEOREM 1. *For a ring R with local units the following are equivalent:*

1) *R is Morita equivalent to a direct sum of local perfect rings.*

2) *R is a direct sum of primary rings satisfying the descending chain condition on finitely generated left ideals.*

3) *R satisfies the descending chain condition on finitely generated left ideals and each idempotent ideal K is a ring theoretic direct summand of R.*

1991 *Mathematics Subject Classifications.* Primary 16P70; Secondary 16D90.
Key words and phrases. Perfect rings, Morita equivalence.

Research partially supported by Hungarian National Foundation for Scientific Research Grant No. 1813.

Akadémiai Kiadó, Budapest

PROOF. The equivalence $1 \Leftrightarrow 2$ is an immediate consequence of Corollary 3.8 in [1].

$2 \Rightarrow 3$. It suffices to show that K is a ring-theoretic direct summand of R if $K^2 = K$. Since $R = \oplus R_i$ and each R_i contains only the trivial idempotent ideals 0 and R_i by the transfinite nilpotency of the radical, the images K_i of K under the projections $R \to R_i$ are R_i or 0. On the other hand, one can deduce easily that K is a direct sum of K_i. Therefore K is a direct summand of R.

$3 \Rightarrow 2$. Clearly, it is enough to see that R is a direct sum of primary rings. Let $\{P_i\}$ be the set of all maximal ideals of R and $Q_i = P_i$ for every index i. Every idempotent of R is obviously a sum of orthogonal primitive idempotents, and for every primitive idempotent $e \in R$ there is a P_i with $e \notin P_i$ and hence $e \notin Q_i$. Therefore $Q_i e = 0$, because Q_i is a ring-theoretic direct summand of R. Consequently, $\bigcap Q_i = \bigcap_{e^2=e \in R} R(1-e) = 0$ holds. Consider now the ring homomorphism

$$\phi : R \to \prod R/Q_i : r \mapsto (\ldots, r + Q_i, \ldots).$$

ϕ is injective, for $\cap Q_i = 0$. Since each idempotent belongs to almost all P_i and hence Q_i, as it is easy to check, and R has local units, ϕ is a monomorphism from R into the direct sum $\oplus R/Q_i$. On the other hand, R/Q_i is obviously a primary ring with radical P_i/Q_i, and $\phi(R) = \oplus R/Q_i$ because Q_i is a direct summand of R. Thus ϕ is an isomorphism between R and $\oplus R/Q_i$, which completes the proof.

A ring R is called *weakly artinian* if its finitely generated left ideals are artinian left R-modules. Let R be a primary weakly artinian ring with local units. By Corollary 3.8 in [1], R is a strongly locally matrix ring over the artinian local ring eRe where $e^2 = e$ is a primitive idempotent in R. Consequently, the radical of R is nilpotent. This fact implies the following assertion.

PROPOSITION 2. *The radical of a weakly artinian primary ring with local units is nilpotent.*

Similarly to the case of artinian rings we have

PROPOSITION 3. *Any finitely generated left ideal of a weakly artinian ring with local units is a noetherian module, i.e. it is of finite length.*

PROOF. Let e be an arbitrary idempotent of R. By the assumption there is a positive integer k with $J^k e = J^n e$ for all $n > k$, consequently $J^k e = 0$ holds by the transfinite nilpotency of the radical J of R. Therefore, similarly to the Hopkin's Theorem on artinian rings, one can deduce that the left R-module Re is of finite length, from which the statement follows because R has local units.

PROPOSITION 4. $\cap J^n = 0$ *holds for the radical* J *of a weakly artinian ring* R *with local units.*

PROOF. By the proof of Proposition 3 for each idempotent $e \in R$ there is an integer n with $J^n e = 0$, i.e., $J^n \subseteq R(1-e)$. Therefore $\cap J^n \subseteq \bigcap_{e^2=e} R(1-e) = 0$, since R has local units.

PROPOSITION 5. *If* P *is any maximal ideal of a weakly artinian ring* R *with local units, then there is an integer* k *with* $P^k = P^n$ *for all* $n > k$.

PROOF. Let $Q = \cap P^n$. By $P + J = Q + J$ we have that P/Q is the radical of the primary ring R/Q. Therefore by Proposition 2 there is an integer k with $(P/Q)^k = 0$, i.e., $P^k = P^n$ for all $n > k$.

Let $\{P_i\}$ be the set of maximal ideals of a weakly artinian ring with local units. Since the idempotents of P_i are trivially contained in $Q_i = \bigcap P_i^n$ and the radical of R is a small ideal, we have $Q_i + Q_j = R$ for all $i \neq j$. Therefore for $i \notin \{i_1, \ldots, i_n\}$ it holds

$$(*) \qquad Q_i + (Q_{i_1} \cap \ldots \cap Q_{i_n}) = R.$$

PROPOSITION 6. *Let* P_i *be the set of maximal ideals in a weakly artinian ring* R *with local units, and* $Q_i = \cap P_i^n$ *for all indices* i. *Then* $\cap Q_i = 0$ *if and only if one of the following conditions is satisfied:*

1) $P_i P_j = P_j P_i$ *for all* i *and* j,
2) $Q_i Q_j = Q_j Q_i$ *for all* i *and* j.

PROOF. Assume indirectly $\cap Q_i \neq 0$. Then there are an element $c \in \cap Q_i$, $c \neq 0$ and an idempotent $e \in R$ with $e = ce = ec$. By Proposition 3 the R-module Re is of finite length, and hence $P_1^{k_1} \ldots P_n^{k_n} Re = 0$ for finitely many maximal ideals $P_1, \ldots, P_n$ and positive integers $k_1, \ldots, k_n$. This shows that $c \notin P_1^{k_1} \ldots P_n^{k_n}$.

1. First assume $P_i P_j = P_j P_i$ for all $i \neq j$. Since $P_i^k + P_j^m = R$ for all $i \neq j$ and integers k, m, we have

$$P_i^k \cap P_j^m = (P_i^k \cap P_j^m)R = (P_i^k \cap P_j^m)(P_i^k + P_j^m) \subseteq P_i^k P_j^m.$$

This ensures $P_i^k \cap P_j^m = P_i^k P_j^m$ for all $i \neq j$. Similarly one can see $P_{i_1}^{m_1} \ldots P_{i_n}^{m_n} = P_{i_1}^{m_1} \cap \ldots \cap P_{i_n}^{m_n}$ for different indices $i_1, \ldots, i_n$. Therefore we get $c \notin P_1^{k_1} \ldots P_n^{k_n} = P_1^{k_1} \cap \ldots \cap P_n^{k_n} \supseteq \cap Q_i$ which contradicts $c \in \cap Q_i$. Thus $\cap Q_i = 0$ holds in this case.

2. Secondly assume $Q_i Q_j = Q_j Q_i$ for all indices i, j. Similarly to the above consideration one can show that $Q_1 \cap \ldots \cap Q_n = Q_1 \ldots Q_n$ for any finite collection $Q_1, \ldots, Q_n$. By $Q_i^2 = Q_i$ we have

$$c \notin P_1^{k_1} \ldots P_n^{k_n} \supseteq Q_1^{k_1} \ldots Q_n^{k_n} = Q_1 \ldots Q_n = Q_1 \cap \ldots \cap Q_n \supseteq \bigcap Q_i,$$

which contradicts $c \in \cap Q_i$. Thus $\cap Q_i = 0$.

Now using the equality $(*)$ we can show $R \tilde{=} \oplus R/Q_i$, where each R/Q_i is clearly a primary ring, in the same way as it was done in the proof of Theorem 1. Observing that a local artinian ring has nonzero one-sided socles, by Theorem 1 and Corollaries 3.6 and 3.8 in [1] we have

THEOREM 7. *Let R be a ring with local units. The following assertions are equivalent:*

1) *R is Morita equivalent to a direct sum of local artinian rings.*

2) *R is a direct sum of primary weakly artinian rings.*

3) *R is a weakly artinian ring and each idempotent ideal K of R is a ring-theoretic direct summand of R.*

4) *R is a weakly artinian ring and any two maximal ideals of R commute.*

5) *R is a weakly artinian ring and any two Q_i commute, where $Q_i = \cap P_i^n$ and the P_i are the maximal ideals of R.*

6) *R is a direct sum of Rees matrix rings over local artinian rings with independent sandwich matrices.*

COROLLARY 8. *A ring R with local units is a direct sum of local artinian rings if and only if it is weakly artinian and satisfies one of the following conditions:*

1) *Any two maximal left ideals commute.*

2) *Any two maximal right ideals commute.*

PROOF. Let R be a weakly artinian ring and P be an arbitrary maximal ideal of R. Then R/P is a simple ring with minimal one-sided ideals and thus R/P is a direct sum of minimal left and right ideals, respectively. From this fact we can see by direct computation that under the assumption of Corollary 8 the maximal ideals of R are maximal one-sided ideals, too. Therefore our statement is an immediate consequence of Theorem 7.

REFERENCES

[1] ANH, P. N., Morita equivalence and tensor product rings, *Comm. Algebra* **17** (1989), 2717–2737. *MR* **91f**:16015

[2] NUMAKURA, K., Theory of compact rings, *Math. J. Okayama Univ.* **5** (1955), 79–93. *MR* **17**–642

[3] NUMAKURA, K., Theory of compact rings. II, *Math. J. Okayama Univ.* **5** (1955), 103–113. *MR* **17**–1223

[3] SZÁSZ, F., Über Ringe mit Minimalbedingung für Hauptrechtsideale, III, *Acta Math. Acad. Sci. Hungar.* **14** (1963), 447–461. *MR* **28**#1211

(Received January 4, 1990)

MTA MATEMATIKAI KUTATÓINTÉZETE
P. O. BOX 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 289–297

ÜBER DIE SYNTHETISCHE BEHANDLUNG DER KRÜMMUNG UND DES SCHMIEGZYKELS DER EBENEN KURVEN IN DER BOLYAI–LOBATSCHEFSKYSCHEN GEOMETRIE

I. VERMES

I

Wir betrachten einen Kurvenbogen $\overset{\frown}{AB}$, der einen konvexen Bereich S mit seiner Sehne AB so begrenzt, daß jede Strecke, die zwei Punkte von $\overset{\frown}{AB}$ verbindet, in Inneren von S läuft. Zu jedem Punkt des Bogens $\overset{\frown}{AB}$ gehört je eine Tangente, die als Grenzlage der zum Berührungspunkt gehörigen Sehnen entsteht.

Setzen wir voraus, daß der Bereich S quadrierbar ist. Der Inhalt von S stimmt mit dem Grenzwert der Inhalte der konvexen Vielecke überein, die durch die Sehne AB und die, mit ihr verbundenen, in den Bogen $\overset{\frown}{AB}$ eingeschriebenen, unbegrenzt verfeinerten, konvexen Streckenzügen begrenzt werden.

Sei der Kurvenbogen $\overset{\frown}{AB}$ rektifizierbar und sei seine Bogenlänge gleich dem Grenzwert der Längen von Polygonzügen, die aus den in $\overset{\frown}{AB}$ eingeschriebenen und unbegrenzt verfeinerten konvexen Streckenzügen bestehen. Wir setzen für das folgende voraus, daß die untersuchten Kurvenbögen eine solche Eigenschaft haben — wie zum Beispiel Kreis-, Horozykel- bzw. Hyperzykelbogen — , nach der der Quotient der Bogenlänge und der zu ihr gehörigen Sehnenlänge gegen 1 strebt, falls die Sehnenlänge gegen 0 geht.

Für die obenerwähnten Kurven können die Begriffe der Krümmung in einem Punkt und der totalen Krümmung eines Bogens $\overset{\frown}{AB}$ (das ist ein Maß für die Richtungsveränderung der Tangenten zwischen A und B) erklärt werden. Da die Richtungsveränderung der Tangenten in der Bolyai–Lobatschefskyschen Geometrie nicht analog der euklidischen Geometrie feststellbar ist, deswegen müßte man den Begriff der Verschiebung auf brauchbare Weise für unsere Geometrie hinüberretten. Es ist daher eine charakteristische Verschiebung der Tangente entlang eines Kurvenbogens zu bestimmen, und dann ist die Richtungsveränderung der Tangente als ein Winkel meßbar.

1991 *Mathematics Subject Classification.* Primary 53A35; Secondary 51K10.

Key words and phrases. Non-Euclidean differential geometry, synthetic differential geometry.

Unterstützt von der Ungarischen Akademie der Wissenschaften im Projekt OTKA Nr. 1615 (1991).

Akadémiai Kiadó, Budapest

Die Verschiebung einer Halbgerade. Die Halbgerade e wird entlang der Strecke AB in die Lage e' verschoben, wenn der Winkel ε zwischen e und AB, und der Winkel ε' zwischen e' und der aus B in die Richtung AB ausgehenden Halbgerade kongruent sind, und alle beide auf derselben Seite von der Geraden AB liegen (Fig. 1).

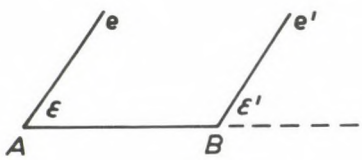


Fig. 1

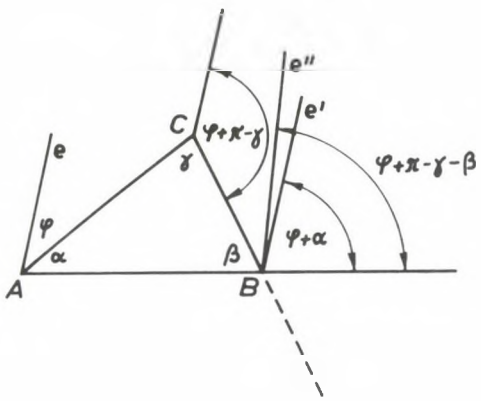


Fig. 2

Falls man die Halbgerade e entlang der Strecke AB bzw. AC und CB in den Punkt B verschiebt, so erhält man die Halbgeraden e' bzw. e'' (Fig. 2). Es ist leicht zu sehen, daß e'' bzw. e' und die Richtung AB den Winkel $\varepsilon'' = \varphi + \pi - \gamma - \beta$ bzw. $\varepsilon' = \varphi + \alpha$ einschließen. Folglich ergibt sich der Winkel von e' und e'' aus

$$\varepsilon'' - \varepsilon' = \pi - (\alpha + \beta + \gamma),$$

der aber der Winkeldefekt des Dreieckes ABC ist. Der Winkeldefekt kann auch auf folgende Weise geschrieben werden:

$$\varepsilon'' - \varepsilon' = \frac{T}{k^2},$$

wobei T der Inhalt des Dreieckes ist, und k die Konstante der Geometrie bedeutet.

Schließt eine Halbgerade f mit e den Winkel ν ein, so schließen beide verschobenen Halbgeraden f' bzw. f'' mit e' bzw. e'' auch den Winkel ν ein, womit das Verschiebungsverfahren in diesem Sinne winkeltreu ist. Falls man eine Halbgerade entlang der Strecke BA bzw. BC und CA aus dem Punkt B in den Punkt A verschiebt, so schließen die erhaltenen Halbgeraden auch den Winkeldefekt von ABC ein.

Auf Grund unserer Kenntnisse für das Dreieck ABC ergibt sich unmittelbar ein Korollar, (das durch vollständige Induktion bewiesen werden kann) wie folgt: Verschieben wir eine Halbgerade e entlang des konvexen Streckenzuges (Polygons), AA_1, A_1A_2, ..., A_{n-1}, B bzw. entlang der Strecke AB in die Lage e'' bzw. e' (Fig. 3), und bezeichne T_{n+1} den Inhalt des konvexen $(n+1)$-Eckes A, A_1, ..., A_{n-1}, B, so kann der Winkel von e'' und e' in der Form $\frac{T_{n+1}}{k^2}$ geschrieben werden.

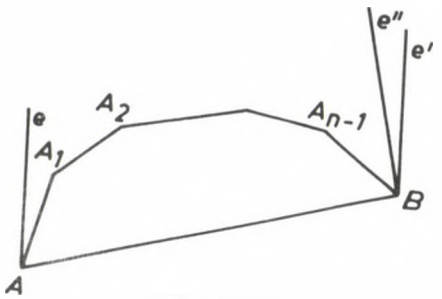


Fig. 3

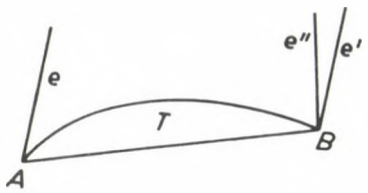


Fig. 4

Die Verschiebung entlang einer Kurve. Es sei ein konvexer Kurvenbogen $\overset{\frown}{AB}$ gegeben, und sei A, A_1, ..., A_{n-1}, B ein konvexes Polygon in $\overset{\frown}{AB}$ eingeschrieben. Das Verschiebungsverfahren entlang eines solchen Polygons gibt den Winkel von e'' und e', dessen Größe $\frac{T_{n+1}}{k^2}$ ist. Verfeinern wir die einbeschriebenen Polygone unbegrenzt, so strebt der Inhalt T_{n+1} zunehmend gegen T, wobei T den Inhalt zwischen dem Bogen $\overset{\frown}{AB}$ und der Strecke AB bedeutet. Die Halbgeraden $e_i''(i = 1, 2, \ldots)$ entfernen sich von e' bis zur Grenzlage e'' (Fig. 4), und gleichzeitig wird der Winkel zwischen e' und e'' zu: $\sphericalangle(e', e'') = \frac{T}{k^2}$. Folglich können wir die Verschiebung einer Halbgeraden e entlang eines konvexen Kurvenbogens so definieren, daß die verschobene Halbgerade e'' mit der die Strecke AB entlang verschobenen Halbgeraden e' den Winkel von der Größe $\frac{T}{k^2}$ einschließt.

DEFINITION. Unter der *totalen Krümmung* des Kurvenbogens $\overset{\frown}{AB}$ verstehen wir den Winkel $\Delta\alpha$, der zwischen den Halbgeraden e und e_B eingeschloßen wird (Fig. 5), wo e_A und e_B die Tangenten der Kurve in den Punkten A bzw. B sind, und e sich durch die Verschiebung von e_A den Bogen $\overset{\frown}{AB}$ entlang ergibt.

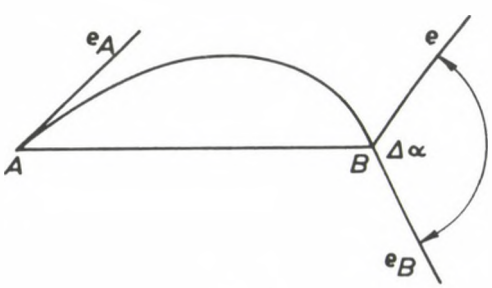


Fig. 5

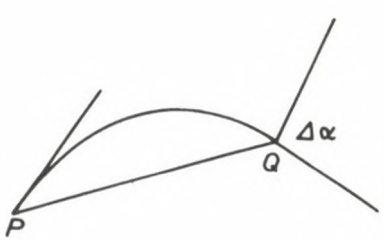


Fig. 6

Die Krümmung einer Kurve in ihrem Punkt P. Betrachten wir einen konvexen Kurvenbogen $\overset{\frown}{PQ}$, dessen totale Krümmung $\Delta\alpha$ und dessen Bogenlänge Δs ist. Die Krümmung dieser Kurve in P ist der Grenzwert des

Quotienten $\frac{\Delta\alpha}{\Delta s}$, wenn der Punkt Q gegen P strebt (d.h. Δs gegen 0) (Fig. 6). Wir setzen im folgenden voraus, daß dieser Grenzwert existiert:

$$K = \lim_{\Delta s \to 0} \frac{\Delta\alpha}{\Delta s}.$$

Wir können voraussetzen, daß K nicht verschwindet.

II

Die Krümmung eines Kreises vom Radius r. Sei der Mittelpunkt eines Kreisbogens $\overset{\frown}{AB}$ mit O und der Zentriwinkel mit Ψ bezeichnet (Fig. 7). Die Tangente in B und die Sehne AB schließen den Winkel β ein. Für die totale Krümmung $\Delta\alpha$ des Bogens $\overset{\frown}{AB}$ erhalten wir

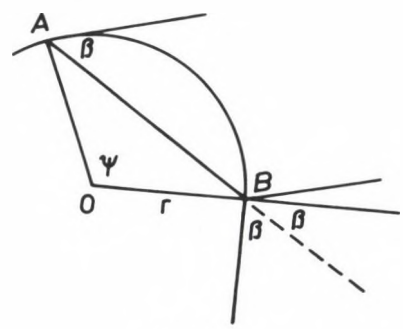


Fig. 7

$$\Delta\alpha = \frac{T}{k^2} + 2\beta,$$

wo T den inhalt des Kreissegmentes bedeutet. Die Bogenlänge von $\overset{\frown}{AB}$ ist Δs[1)]

$$\Delta s = k\Psi \operatorname{sh}\frac{r}{k}.$$

Der Inhalt des Sektors OAB[2)] ist:

$$T_1 = k^2\Psi(\operatorname{ch}\frac{r}{k} - 1),$$

und der Inhalt des Dreieckes OAB ergibt sich zu

$$T_2 = k^2\left\{\pi - \left[\Psi + 2\left(\frac{\pi}{2} - \beta\right)\right]\right\} = k^2(2\beta - \Psi).$$

Damit ist der Inhalt T des Kreissegmentes

1) Vgl. [3] S. 89, [4] S. 99, [5] S. 184
2) Vgl. [3] S. 95, [4] S. 101, [5] S. 243

$$T = T_1 - T_2 = k^2 \left(\Psi \mathrm{ch} \frac{r}{k} - 2\beta \right),$$

und es gilt

$$\Delta\alpha = \Psi \mathrm{ch} \frac{r}{k}.$$

Insgesamt erhalten wir

$$\frac{\Delta\alpha}{\Delta s} = \frac{\Psi \mathrm{ch} \frac{r}{k}}{k\Psi \mathrm{sh} \frac{r}{k}} = \frac{1}{k} \mathrm{cth} \frac{r}{k}.$$

Wenn Δs gegen 0 strebt (d. h. $\Psi \to 0$), bleibt der Quotient $\frac{\Delta\alpha}{\Delta s}$ unverändert, weil er von Ψ unabhängig ist. Die Krümmung $K(r)$ des Kreises vom Radius r ist damit:

$$K(r) = \frac{1}{k} \mathrm{cth} \frac{r}{k}.$$

Die Krümmung des Horozykels (Grenzkreises). Betrachten wir den Horozykelbogen $\overset{\frown}{AB}$. Bezeichne $2x = AB$ und $\Pi(x)$ den zum Lote x gehörigen Parallelwinkel. Die Tangente in A und die Sehne AB schließen den Winkel β ein (Fig. 8):

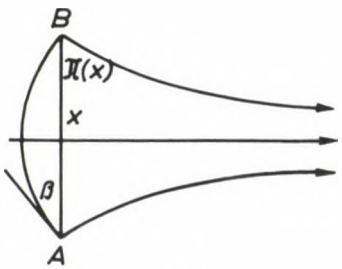


Fig. 8

$$\beta = \frac{\pi}{2} - \Pi(x).$$

Der Inhalt T_1 des zur Sehne AB gehörigen Grenzkreissektors bzw. die Bogenlänge $\overset{\frown}{AB}$ können auf folgende Weise geschrieben werden:

$$T_1 = k\overset{\frown}{AB} \quad ^{3)}$$

bzw.

$$\overset{\frown}{AB} = 2k\mathrm{sh} \frac{x}{k}. \quad ^{4)}$$

---

[3] Vgl. [1] bzw. in [2] Bolyai § 32. V. S. 32.

[4] Vgl. [4] S. 106, [5] S. 187.

Folglich ist der Inhalt T des Grenzkreissegmentes:

$$T = 2k^2 \mathrm{sh}\frac{x}{k} - 2k^2\beta.$$

Die totale Krümmung von $\overset{\frown}{AB}$ ist:

$$\Delta\alpha = \frac{1}{k^2}T + 2\beta = 2\mathrm{sh}\frac{x}{k}.$$

Weil $\Delta s = 2k\mathrm{sh}\frac{x}{k}$ besteht, so gibt sowohl der Quotient $\frac{\Delta\alpha}{\Delta s}$ als auch sein Grenzwert (im Falle $\Delta s \to 0$) den Wert $\frac{1}{k}$. Die Krümmung des Horozykels (in allen Punkten) wird $\frac{1}{k}$.

Die Krümmung eines Hyperzykels (einer Abstandslinie) vom Abstand l. Zum Hyperzykelbogen $\overset{\frown}{AB}$ gehört eine Strecke $A_1B_1 = x$ auf der Grundlinie, und die Tangente in A schließt den Winkel β mit der Sehne AB ein, wobei $\vartheta + \beta = \frac{\pi}{2}$ besteht (Fig. 9). Die Bogenlänge Δs des Hyperzykelbogens $\overset{\frown}{AB}$ ist

$$\Delta s = x\mathrm{ch}\frac{l}{k}. \quad [5]$$

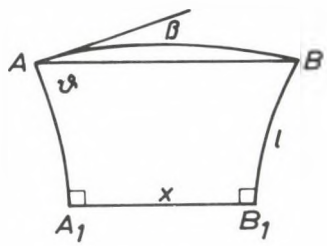


Fig. 9

Als Inhalt T_1 zwischen dem Bogen $\overset{\frown}{AB}$ und der zu ihm gehörigen Strecke A_1B_1 erhalten wir

$$T_1 = kx\mathrm{sh}\frac{l}{k}. \quad [6]$$

Der Inhalt T_2 des Saccheri-schen Viereckes ABB_1A_1 ist

$$T_2 = k^2(\pi - 2\vartheta).$$

Also ergibt sich der Inhalt T des Hyperzykelsegmentes:

$$T = T_1 - T_2 = k^2\left(\frac{x}{k}\mathrm{sh}\frac{l}{k} - 2\beta\right).$$

---

[5] Vgl. [3] S. 95, [4] S. 119, [5] S. 184.

[6] Vgl. [3] S. 95, [4] S. 120, [5] S. 243.

Die totale Krümmung $\Delta\alpha$ von $\overset{\frown}{AB}$ ist:

$$\Delta\alpha = \frac{T}{k^2} + 2\beta = \frac{x}{k}\operatorname{sh}\frac{l}{k}.$$

Daraus folgt sofort, daß der Wert und der Grenzwert (im Falle $\Delta s \to 0$) von $\frac{\Delta\alpha}{\Delta s}$ übereinstimmen. Es ergibt sich für die Krümmung $K(l)$ eines Hyperzykels vom Abstand l (in allen Punkten):

$$K(l) = \frac{1}{k}\operatorname{th}\frac{l}{k}.$$

Die Figur 10 zeigt die graphische Darstellung der Krümmungen der Kreise, des Horozykels und der Hyperzykeln.

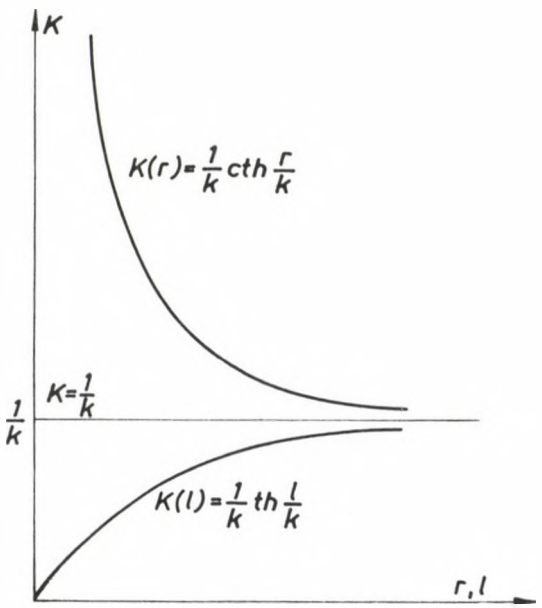


Fig. 10

Der Schmiegzykel ebener Kurven. Betrachten wir eine Kurve g und ihre Punkte P, Q. Die Normale in P und die die Strecke PQ senkrecht halbierende Gerade bestimmen ein Strahlbüschel in der Ebene. Sei c der Zykel durch P, Q, der zu diesem Büschel gehört (Fig. 11). Dieser Zykel berührt die Kurve g. Setzen wir voraus, daß der Zykel c eine Grenzlage hat, falls Q gegen P strebt. Diese Grenzlage wird *Schmiegzykel* der Kurve g im Punkt P genannt.

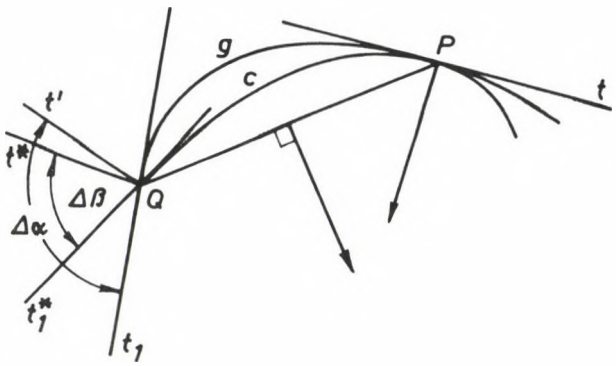


Fig. 11

Wir beweisen, daß die Krümmung der Kurve g und die Krümmung des Schmiegzykels übereinstimmen.

Verschieben wir die zu P gehörige gemeinsame Tangente entlang der Kurve g bzw. des Zykels c in den Punkt Q: sind die erhaltenen Halbgeraden t' bzw. t^*. Bezeichnen t_1 bzw. t_1^* die Tangenten von g bzw. c im Punkt Q. Es seien $\sphericalangle(t',\ t_1) = \Delta\alpha,\ \sphericalangle(t^*,\ t_1^*) = \Delta\beta$, und seien die Längen der Bogen $\overset{\frown}{PQ}$ von g bzw. c mit Δs bzw, $\Delta\sigma$ bezeichnet. Falls Q gegen P strebt, so gilt $\Delta s \to 0$ und $\Delta\sigma \to 0$, und auch der Winkel von t_1 und t_1^* strebt gegen 0, weil sowohl der Winkel von t_1 und PQ als auch der Winkel von t_1^* und PQ gegen 0 strebt. Wenn $Q \to P$ strebt, so strebt sowohl der Inhalt zwischen der Sehne PQ und der Kurve g als auch der Inhalt zwischen der Sehne PQ und dem Zykel c gegen 0, folglich wird der Inhalt zwischen g und c, und der Winkel von t' und t^* beliebig klein. Daraus folgt, daß die Abweichung zwischen $\Delta\alpha$ und $\Delta\beta$ beliebig klein wird, und so $\frac{\Delta\alpha}{\Delta\beta}$ gegen 1 strebt:

$$\lim_{Q\to P} \frac{\Delta\alpha}{\Delta\beta} = 1.$$

Weil

$$\frac{\frac{\Delta\alpha}{\Delta s}}{\frac{\Delta\beta}{\Delta\sigma}} = \frac{\Delta\alpha}{\Delta\beta}\frac{PQ}{\Delta s}\frac{\Delta\sigma}{PQ},$$

und $\lim\limits_{Q\to P} \frac{\Delta s}{PQ} = 1$ bzw. $\lim\limits_{Q\to P} \frac{\Delta\sigma}{PQ} = 1$ bestehen, so ist

$$\lim_{Q\to P} \frac{\frac{\Delta\alpha}{\Delta s}}{\frac{\Delta\beta}{\Delta\sigma}} = \frac{K_g}{K_c} = 1,$$

wo K_g bzw. K_c die Krümmung von g bzw. c in P bedeutet. Also ist $K_g = K_c$, was zu beweisen war.

LITERATURVERZEICHNIS

[1] BOLYAI, J., *Appendix Scientiam spatii absolute veram exhibens: a veritate aut falsitate axiomatis XI Euclidei a priori haud unquam decidenda, independentem: adiecta ad casum falsitatis, quadratura circuli geometrica*, Akadémiai Kiadó, Budapest, 1952.

[2] BONOLA, R., *Non-Euclidean geomctry, a critical and historical study of its developments*, With a Supplement containing the George Bruce Halsted translations of "The science of absolute space" by John Bolyai and "The theory of parallels" by Nicholas Lobachevski, Dover Publications, Inc., New York, 1955. *MR* **16**-1145

[3] LIEBMANN, H., *Nichteuklidische Geometrie*, Zweite neubearbeitete Auflage, Sammlung Schubert, No. 49, G. J. Göschen'sche Verlagshandlung G.m.b.H., Berlin und Leipzig, 1912. *Jb. Fortschritte Math.* **43**, 557

[4] PERRON O., *Nichteuklidische Elementargeometrie der Ebene*, Mathematische Leitfäden, B. G. Teubner Verlagsgesellschaft, Stuttgart, 1962 *MR* **25** # 2489

[5] SZÁSZ, P., *Bevezetés a Bolyai–Lobacsevszkij-féle geometriába* [Introduction to Bolyai–Lobačevskian geometry], Disquisitiones Mathematicae Hungaricae, No. 5, Akadémiai Kiadó, Budapest, 1973. *MR* **54** # 1072

(Eingegangen am 1. Februar 1990.)

BUDAPESTI MŰSZAKI EGYETEM
GÉPÉSZMÉRNÖKI KAR
GEOMETRIA TANSZÉK
EGRI JÓZSEF U.1
H–1521 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 299–316

SATURATION ORDERS OF SOME APPROXIMATION PROCESSES IN CERTAIN BANACH SPACES

S. P. YADAV

1. Introduction

1.1 *Some Banach spaces of functions.* We write X, to mean the space $C[-1,1]$ of all continuous functions, which is a Banach space if the norm is given by

$$\|f\|_{X=C} = \max_{-1\leqq x\leqq 1} |f(x)| \tag{1.1.1}$$

or the spaces L^p $(1 \leqq p < \infty)$ with the weight function

$$P^{(\alpha,\beta)}(\theta) = (\sin\theta/2)^{2\alpha+1}(\cos\theta/2)^{2\beta+1}, \quad (\alpha \geqq \beta \geqq -1/2) \tag{1.1.2}$$

$(x = \cos\theta)$, which are Banach spaces if endowed with the norms

$$\|f\|_{X=L^p} = \left\{ \int_0^\pi |f(\cos\theta)|^p P^{(\alpha,\beta)}(\theta) d\theta \right\}^{1/p}. \tag{1.1.3}$$

Again, $X = L^\infty$ is a Banach space of functions if the norm is given by

$$\|f\|_\infty = \operatorname{ess\,sup}_{0\leqq\theta\leqq\pi} |f(\cos\theta)|, \quad (x = \cos\theta) \tag{1.1.4}$$

and also $X = M$ the space of all regular finite Borel measures on $[-1,1]$ is Banach space with the norm

$$\|\mu\|_M = \int_0^\pi |d\mu(\cos\theta)|. \tag{1.1.5}$$

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 41A40; Secondary 41A65, 41A25.

Key words and phrases. Saturation orders, approximation method, Jacobi polynomials, convolution, generalized translates.

Akadémiai Kiadó, Budapest

With $f \in X$ $(= C$ or L^p $(1 \leqq p < \infty))$ we associate the Fourier–Jacobi expansion

$$f(\cos\theta) \sim \sum_{n=0}^{\infty} \hat{f}(n)\omega_n^{(\alpha,\beta)} R_n^{(\alpha,\beta)}(\cos\theta), \tag{1.1.6}$$

where the Fourier–Jacobi transforms $\hat{f}(n)$ are defined by

$$\hat{f}(n) = \int_0^{\pi} f(\cos\theta) R_n^{(\alpha,\beta)}(\cos\theta) P^{(\alpha,\beta)}(\theta) d\theta \tag{1.1.7}$$

$$R_n^{(\alpha,\beta)}(\cos\theta) \stackrel{\text{def}}{=} P_n^{(\alpha,\beta)}(\cos\theta) \,/\, P_n^{(\alpha,\beta)}(1)$$

such that

$$\int_0^{\pi} R_n^{(\alpha,\beta)}(\cos\theta) R_m^{(\alpha,\beta)}(\cos\theta) P^{(\alpha,\beta)}(\theta) d\theta = \delta_{nm}\{\omega_n^{(\alpha,\beta)}\}^{-1}, \tag{1.1.8}$$

δ_{nm} being the Kronecker delta and $P_n^{(\alpha,\beta)}(\cos\theta)$ the Jacobi polynomial of degree n and order (α,β) (see Szegő [7]) and

$$\begin{aligned} \omega_n^{(\alpha,\beta)} &= \frac{(2n+\alpha+\beta+1)\Gamma(n+\alpha+\beta+1)\Gamma(n+\alpha+1)}{\Gamma(n+\beta+1)\Gamma(n+1)\Gamma(\alpha+1)\Gamma(\alpha+1)} \\ &= \frac{n^{2\alpha+1}}{[\Gamma(\alpha+1)]^2}(1+O(1/n)) \stackrel{\text{def}}{=} n^{2\alpha+1}L(n) \end{aligned} \tag{1.1.9}$$

(definition of $L(n)$).

1.2 *Convolution structure for certain matrix transforms.* Generalized translate of f with expansion (1.1.6) is defined (see [1]) as $T_\phi f$ with expansion

$$T_\phi f(\cos\theta) \sim \sum_{n=0}^{\infty} \hat{f}(n)\omega_n^{(\alpha,\beta)} R_n^{(\alpha,\beta)}(\cos\theta) R_n^{(\alpha,\beta)}(\cos\phi). \tag{1.2.1}$$

It is known that T_ϕ is a positive operator for $\alpha \geqq \beta \geqq -1/2$ and has operator norm 1 (see [6]). Convolution structure introduced by Askey and Wainger [1] is given as

$$(f_1 * f_2)(\cos\theta) = \int_0^{\pi} T_\phi f_1(\cos\theta) f_2(\cos\phi) P^{(\alpha,\beta)}(\phi)\, d\phi \tag{1.2.2}$$

where the binary operation $*$ (called convolution) is commutative, associative

$$\|f * g\|_X \leqq \|f\|_1 \|g\|_X, \quad f \in L^1, g \in X \tag{1.2.3}$$

and

$$(f * g)\hat{}(n) = f\hat{}(n)g\hat{}(n); \quad f, g \in X. \tag{1.2.4}$$

Also we have (see [2] and [6])

$$R_n^{(\alpha,\beta)}(\cos\theta)R_n^{(\alpha,\beta)}(\cos\phi) =$$

$$= \int_0^\pi R_n^{(\alpha,\beta)}(\cos\psi)K(\cos\theta, \cos\phi, \cos\psi)P^{(\alpha,\beta)}(\psi)\,d\psi, \tag{1.2.5}$$

where $K(\cos\theta, \cos\phi, \cos\psi)$ is a symmetric function in θ, ϕ, ψ and is positive for $\alpha \geqq \beta \geqq -1/2$, and

$$\int_0^\pi K(\cos\theta, \cos\phi, \cos\psi)P^{(\alpha,\beta)}(\psi)\,d\psi = 1. \tag{1.2.6}$$

Thus the generalized translate of $f(\cos\theta) \in L^1$ has the form

$$T_\phi f(\cos\theta) \stackrel{\text{def}}{=} f(\cos\theta, \cos\phi) =$$

$$= \int_0^\pi f(\cos\psi)K(\cos\theta, \cos\phi, \cos\psi)P^{(\alpha,\beta)}(\psi)\,d\psi \tag{1.2.7}$$

with the Fourier–Jacobi expansion given by (1.1.6). If we denote the partial sum of (1.1.6) by $s_n(f, \cos\theta, X)$ then by (1.2.5) and (1.2.6),

$$s_n(f, \cos\theta, X) - f(\cos\theta) =$$

$$(1.2.8) = \sum_{\gamma=0}^{n} \int_0^\pi (f(\cos\phi) - f(\cos\theta))\omega_\gamma^{(\alpha,\beta)} R_\gamma^{(\alpha,\beta)}(\cos\theta)R_\gamma^{(\alpha,\beta)}(\cos\phi)P^{(\alpha,\beta)}(\phi)d\phi =$$

$$= \int_0^\pi [T_\psi f(\cos\theta) - f(\cos\theta)]L_n R_n^{(\alpha+1,\beta)}(\cos\psi)P^{(\alpha,\beta)}(\psi)\,d\psi$$

(see [8]), where

$$L_n \stackrel{\text{def}}{=} \frac{\Gamma(n+\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(n+\beta+1)} P_n^{(\alpha+1,\beta)}(1) = \frac{\alpha+1}{2n+\alpha+\beta+2}\omega_n^{(\alpha+1,\beta)} \tag{1.2.9}$$

$$\sim n^{2\alpha+2}\{1 + O(n^{-1})\} \stackrel{\text{def}}{=} n^{2\alpha+2}L_n$$

so that L_n is positive slowly varying function of n (see Bavinck [2]). We consider matrices $((\lambda_{n,k}))$ with $\lambda_{n,0}=1$ or $((\Delta\lambda_{n,k}))$, $\Delta\lambda_{n,k}=\lambda_{n,k}-\lambda_{n,k+1}$, $\sum_{k=0}^{n}\Delta\lambda_{n,k}=1$, attributing them *primary* and *secondary* matrices, respectively. Mapping $\Lambda\colon\{s_n\}\to\{\sigma_n\}$ being linear and an endomorphism in a sequence space is a transformation from X to X given by

$$\sigma_n^{(\Lambda)}(f,\cos\theta,X)=\sum_{k=0}^{n}\Delta\lambda_{n,k}\,s_k(f,\cos\theta,X).$$

Thus, by (1.2.7),

$$\begin{aligned}&\sigma_n^{(\Lambda)}(f,\cos\theta,X)-f(\cos\theta)=\\ &=\int_0^{\pi}[T_\psi f(\cos\theta)-f(\cos\theta)]K_n^{(\Lambda)}(\cos\psi)P^{(\alpha,\beta)}(\psi)\,d\psi=\\ &=(f*K_n^{(\Lambda)})(\cos\theta)-f(\cos\theta),\end{aligned}\tag{1.2.10}$$

where

$$K_n^{(\Lambda)}(\cos\psi)\overset{\text{def}}{=}K_n^{(\Lambda)}(\psi)=\sum_{k=0}^{n}\Delta\lambda_{n,k}L_kR_k^{(\alpha+1,\beta)}(\cos\psi)\tag{1.2.11}$$

(for details see [8]).

It may be remarked that primary and secondary lower triangular matrices $((\lambda_{n,k}))$ and $((\Delta\lambda_{n,k}))$ are interconnected in the sense that any concrete example of one reveals the other. Most of the summability kernels and approximation kernels are defined by $((\Delta\lambda_{n,k}))$. In general, for the matrix $((\lambda_{n,k}))$ its (n,k)-th element is

$$\lambda_{n,k}=\begin{cases}1-\sum_{\nu=0}^{k-1}\Delta\lambda_{n,\nu}, & k\leqq n,\\ 0, & k>n.\end{cases}\tag{1.2.12}$$

Some interesting and commonly known matrices are given by

$$\text{(i)}\quad\Delta\lambda_{n,k}=\begin{cases}1/(n+1), & k\leqq n\\ 0, & k>n,\end{cases}\qquad\text{(used for } (C,1)\text{-mean)}\tag{1.2.13}$$

$$\text{(ii)}\quad\Delta\lambda_{n,k}=\begin{cases}A_{n-k}^{\mu-1}/A_n^{\mu}, & k\leqq n\\ 0, & k>n,\end{cases}\qquad\text{(used for } (C,\mu)\text{-mean)}\tag{1.2.14}$$

(1.2.15) (iii) $\Delta\lambda_{n,k} = \begin{cases} p_{n-k}/P_n, & k \leqq n \\ 0, & k > n, \end{cases}$ (used for (N, P_n)-mean)

$$P_n = \sum_{\nu=0}^{n} p_\nu \neq 0 \quad \text{are real or complex numbers.}$$

(1.2.16)

(iv) $\Delta\lambda_{n,k} = \begin{cases} \dfrac{1}{(k+1)\sum_{j=0}^{n} \frac{1}{j+1}}, & k \leqq n \\ 0, & k > n, \end{cases}$ (used for $(R, \log n, 1)$-mean)

(1.2.17) (v) $\Delta\lambda_{n,k} = \begin{cases} 1, & k = n \\ 0, & k \neq n, \end{cases}$ (used for identity transformation)

and there are many matrices which fall under the generality of our work. Some other examples are

(1.2.18)

(vi) $\Delta\lambda_{n,k} = \begin{cases} \frac{1}{(n+1)}\left(2 - (n+1)\Big/(k+1)\sum_{j=0}^{n} 1/(j+1)\right), & k \leqq n \\ 0, & k > n, \end{cases}$

and

(1.2.19) (vii) $\Delta\lambda_{n,k} = \begin{cases} \frac{1}{(n-k+1)}\Big/\sum_{j=0}^{n} \frac{1}{(j+1)}, & k \leqq n \\ 0, & k > n. \end{cases}$

We notice that matrices $((\Delta\lambda_{n,k}))$ whose (n,k)-th elements are given by (1.2.13) to (1.2.20) are such that $\lambda_{n,0} = \sum_{k=0}^{n} \Delta\lambda_{n,k} = 1$. Thus their corresponding primary forms $((\lambda_{n,k}))$ are known by (1.2.12) which we use somewhere else for the characterization of functions which allow certain known orders.

1.3 *Approximation processes on Banach spaces.* Modulus of continuity in a Banach space X through the concept of generalized translates has been adjudged by Bavinck [2] as

(1.3.1) $$\omega(\phi, f, X) \overset{\text{def}}{=} \omega(\phi) = \sup_{0 \leqq \psi \leqq \phi} \|T_\psi f(\cdot) - f(\cdot)\|_X.$$

Again if $c \in R^+$ and

(1.3.2) $$\omega(\phi, f, X) \leqq c\phi^r$$

then f is said to belong a Lipschitz space $\mathrm{Lip}(r, X)$, $(0 < r \leqq 2)$. If

(1.3.3) $$\|f\|_{\mathrm{Lip}(r,X)} = \|f\|_X + \sup_{n \in Z^+} (n^r \omega(n^{-1}, f, X))$$

is a norm in $\mathrm{Lip}(r, X)$ then this is a Banach space and $\mathrm{Lip}(r, X) \subseteq X$. Recently, the author [8] gave the following results.

THEOREM A. *Let* $((\Delta\lambda_{n,k}))$ *be a lower triangular matrix such that* $\sum_{k=0}^{n} \Delta\lambda_{n,k} = 1$. *Let* $\{\Delta\lambda_{n,k}\}$ *be non-negative and non-decreasing with respect to* k *and*

$$\frac{\Delta\lambda_{n,k}}{\Delta\lambda_{n,i}} \leqq A, \quad i \leqq k \leqq n. \tag{1.3.4}$$

Then the saturation order for Λ*-transform process of approximation to* $f \in X$ *through* (1.1.6) *is given by*

$$\|f(\cdot) - \sigma_n^{(\Lambda)}(f,(\cdot),X)\|_X \equiv \|f(\cdot) - (f * K_n^{(\Lambda)})(\cdot)\|_X \leqq$$

$$\leqq A\Big(\sum_{k=0}^{n} \frac{\omega(1/(k+1))}{(k+1)^{3/2+\alpha}} \sum_{\nu=0}^{k} \Delta\lambda_{n,n-\nu}(n-\nu+1)^{\alpha+1/2} L(n-\nu) + n^{-2\beta-2}\Big) \tag{1.3.5}$$

$$(\alpha \geqq \beta \geqq -1/2)$$

and Favard's class or the saturation class $F(X,\sigma^{(\Lambda)})$ *is a collection of all* $f \in X$ *for which the right-hand side tends to zero as* $n \to \infty$.

Since the non-negative and non-decreasing nature of $\{\Delta\lambda_{n,k}\}$ with $\sum_{k=0}^{n} \Delta\lambda_{n,k} = 1$ implies $\Delta\lambda_{n,k} \leqq 1/(n-k+1)$, so we can avoid condition (1.3.4), which is used in the proof of Q_1 only (see [8]). Thus a modified form of Theorem A is

THEOREM A1. *Let* $((\Delta\lambda_{n,k}))$ *be as in Theorem A. Then the* Λ*-transform process of approximation is saturated with the order given by*

$$\|f(\cdot) - \sigma_n^{(\Lambda)}(f,(\cdot),X)\|_X \equiv \|f(\cdot) - (f * K_n^{(\Lambda)})(\cdot)\|_X \leqq$$

$$\leqq A\Big(\sum_{k=0}^{n} \frac{\omega(1/(k+1))}{(k+1)^{\alpha+3/2}} \sum_{\nu=0}^{k} \Delta\lambda_{n,n-\nu}(n-\nu+1)^{\alpha+1/2} L(n-\nu) + n^{-1}\Big) \tag{1.3.6}$$

and the saturation class or Favard's class $F(X,\sigma^{(\Lambda)})$ *is the collection of all* $f \in X$ *for which the right-hand side of* (1.3.6) *tends to zero as* $n \to \infty$.

Also $\sum_{k=0}^{n} \Delta\lambda_{n,k} = 1$. Thus the series $\sum_{k=0}^{\infty} \Delta\lambda_{n,k}$ is of non-negative terms and converges. Consequently, by Pringsheim's theorem, for every $\varepsilon > 0$ there exists a value of k, say n_0, such that $n_0 \Delta\lambda_{n,n_0} < \varepsilon$, so that we have $Q_1 < A\varepsilon$ (see the proof of Q_1 in [8]). Thus a more refined form of Theorem A1 is

THEOREM A2. *Let* $((\Delta\lambda_{n,k}))$ *be a lower triangular matrix with* $\sum_{k=0}^{n}\Delta\lambda_{n,k}=1$. *Let* $\{(k+1)^{\alpha+1/2}\Delta\lambda_{n,k}\}$ *be non-negative and non-decreasing. Then the saturation order for* Λ*-process of approximation in* X *through* (1.1.6) *is given by*

$$\|f(\cdot)-\sigma_n^{(\Lambda)}(f,(\cdot),X)\|_X\equiv\|f(\cdot)-(f*K_n^{(\Lambda)})(\cdot)\|_X\leqq$$

$$\leqq A\Big(\sum_{k=0}^{n}\frac{\omega(1/(k+1))}{(k+1)^{3/2+\alpha}}\sum_{\nu=0}^{k}\Delta\lambda_{n,n-\nu}(n-\nu+1)^{\alpha+1/2}L(n-\nu)+n^{-2\beta-2}\Big). \tag{1.3.7}$$

The saturation class $F(X,\sigma^{(\Lambda)})$ *is the collection of all* $f\in X$ *for which the right-hand side of* (1.3.7) *tends to zero as* $n\to\infty$.

Again if we restrict the modulus of continuity by

$$\omega\Big(\frac{1}{k+1}\Big)\geqq\begin{cases}A(k+1)^{\alpha-2\beta-3/2}(n-k+1)^{-\alpha-1/2}\\ \text{or}\\ B(k+1)^{\alpha-2\beta-1/2}(n-k+1)^{-\alpha-3/2}\end{cases} \tag{1.3.8}$$

(n large enough, $k=0,1,2,\ldots$; $\alpha\geqq\beta\geqq-1/2$), then we have

THEOREM B. *Let* $\{\Delta\lambda_{n,k}\}$ *be given as in Theorem A and let* (1.3.8) *be satisfied. Then*

$$\|f(\cdot)-\sigma_n^{(\Lambda)}(f,(\cdot),X)\|_X\leqq$$

$$\leqq A\Big(\sum_{k=0}^{n}\frac{\omega(1/(k+1))}{(k+1)^{3/2+\alpha}}\sum_{\nu=0}^{k}\Delta\lambda_{n,n-\nu}(n-\nu+1)^{\alpha+1/2}L(n-\nu)\Big) \tag{1.3.9}$$

and the saturation classes are given as in the previous cases.

REMARK 1.

$$\int_0^{\pi}|K_n^{(\Lambda)}(\theta)|P^{(\alpha,\beta)}(\theta)\,d\theta\geqq An\log n\,(\Delta\lambda_{n,n})\times$$

$$\times\int_0^{\pi}\Big|\sum_{\eta'}^{\eta}k^{2\alpha+1}(\log k)^{-1}L(k)R_k^{(\alpha+1,\beta)}(\cos\theta)\Big|P^{(\alpha,\beta)}(\theta)\,d\theta$$

$$>An\log n\,(\Delta\lambda_{n,n}),\quad(0\leqq\eta'<\eta\leqq n).$$

(Appearance of η, η' depend upon $\Delta\lambda_{n,k}$. Thus the choice of $((\Delta\lambda_{n,k}))$ may

make the integral on the right-hand side to have a lower bound $B > 0$.)

$$\geqq An^{-\alpha+1/2} n^{\alpha+1/2} \log n \, (\Delta\lambda_{n,n}) \geqq$$

$$(1.3.10) \geqq A \sum_{k=0}^{n} (k+1)^{-\alpha+3/2} \sum_{\nu=0}^{k} \Delta\lambda_{n,n-\nu} (n-\nu+1)^{\alpha+1/2} L(n-\nu) >$$

$$> A \sum_{k=0}^{n} \frac{\omega(1/(k+1))}{(k+1)^{\alpha+3/2}} \sum_{\nu=0}^{k} \Delta\lambda_{n,n-\nu} (n-\nu+1)^{\alpha+1/2} L(n-\nu)$$

(for some $f \in X$), i.e. there exists non-trivial element for which the inequalities are reversed. This exhibits the saturation property (see [5], p. 88) if the order tends to zero as $n \to \infty$.

The following corollary of Theorem A is noteworthy. If we denote $(C, 1)$-mean of (1.1.6) by $S_n^1(f, \cos\theta, X)$ then the $(C, 1)$-transform defined by the matrix (1.2.13) has the following estimate:

COROLLARY 1. *For the series* (1.1.6), $\alpha \geqq \beta \geqq -1/2$

$$(1.3.11) \qquad \|f(\cdot) - S_n^1(f, (\cdot), X)\|_X \leqq An^{\alpha-1/2} \sum_{k=0}^{n} \frac{\omega(1/(k+1))}{(k+1)^{1/2+\alpha}}$$

(here we write $n^{-\alpha-2\beta-3/2} = O(1)$ *and notice that* (1.3.4) *is satisfied).*

Condition (1.3.4) is necessary to avoid identity transformation ($\Delta\lambda_{n,k} = 0$, $k \neq n$; $= 1, k = n$), for

$$(1.3.12) \qquad \frac{\Delta\lambda_{n,k}}{\Delta\lambda_{n,i}} = \frac{0}{0}, \quad i \leqq k \leqq n$$

is unwanted which arises in Theorems A and B. However, this identity transformation is applicable in A1, A2 though then we get only estimates for the partial sums. Also the restriction (1.3.4) is not quite superfluous as the proof suggests.

From Theorem A1 or A2 we conclude by substituting $\Delta\lambda_{n,k} = 1$ for $n = k$ and zero otherwise

$$(1.3.13) \qquad \|S_n(f, (\cdot), X)\|_X \leqq \begin{cases} An^{\alpha+1/2}, \alpha > -1/2, \\ A \log n, \alpha = -1/2, \end{cases}$$

where $S_n(f, (\cdot), X)$ is the partial sum of (1.1.6) associated with any $f \in X$.

Our other theorem in [8] along with Theorems A and B is also interesting one. If we consider a lower triangular matrix $((\Delta\lambda_{n,k}))$ with $\lambda_{n,0} \equiv \equiv \sum_{k=0}^{n} \Delta\lambda_{n,k} = 1$ then, summing by parts, using $\lambda_{n,n+1} = \lambda_{n,n+2} = \ldots = 0$, we

have, from (1.2.10)

$$\sigma_n^{(\Lambda)}(f,\cos\theta,X)-f(\cos\theta)=$$

(1.3.14)

$$=\sum_{k=0}^{n}\Delta^2\lambda_{n,k}(k+1)\{S_k^1(f,\cos\theta,X)-f(\cos\theta)\},$$

and (1.3.11) yields

THEOREM C. *We have, for* $\alpha\geqq\beta\geqq-1/2$

(1.3.15)

$$\|f(\cdot)-\sigma_n^{(\Lambda)}(f,(\cdot),X)\|_X\leqq A\Big(\sum_{k=0}^{n}|\Delta^2\lambda_{n,k}|k^{\alpha+1/2}\sum_{\nu=0}^{k}\frac{\omega(1/(\nu+1))}{(\nu+1)^{1/2+\alpha}}\Big).$$

2. Some new processes of approximation

2.1 *Statement of results to be proved.* We have in mind that the structure given by (1.2.10) defines approximation processes for an $f\in X$ when $\{\Delta\lambda_{n,k}\}$ or $\{(k+1)^{\alpha+1/2}\Delta\lambda_{n,k}\}$ are non-negative and non-decreasing. Here we are interested in probing the other aspects, too, i.e. what does it happen when $\{(k+1)^{\alpha+1/2}\Delta\lambda_{n,k}\}$ is non-negative and non-increasing for $0\leqq k\leqq n$? To this end we prove

THEOREM 1. *Let* $((\Delta\lambda_{n,k}))$ *be a lower triangular matrix such that* $\sum_{\nu=0}^{n}\Delta\lambda_{n,\nu}=1$ *and let* $\{(k+1)^{\alpha+1/2}\Delta\lambda_{n,k}\}$ *be non-negative and non-increasing for* $0\leqq k\leqq n$ *and*

$$\frac{\Delta\lambda_{n,k}}{\Delta\lambda_{n,i}}\leqq An^{\alpha+1/2},\quad (i,k)\leqq n, \tag{2.1.1}$$

Then the saturation order for the Λ*-transform process of approximation to* $f\in X$ *through* (1.1.6) *is given by*

$$\|f(\cdot)-\sigma_n^{(\Lambda)}(f,(\cdot),X)\|_X\leqq$$

$$\leqq A\Big(\sum_{k=0}^{n}\frac{\omega(1/(k+1))}{(k+1)^{3/2+\alpha}}\sum_{\nu=0}^{k}(\nu+1)^{\alpha+1/2}\Delta\lambda_{n,\nu}L(\nu)+n^{-2\beta-2}\Big) \tag{2.1.2}$$

for $\alpha\geqq\beta\geqq-1/2$. *The saturation class or Favard's class* $F(X,\sigma^{(\Lambda)})$ *is the collection of all* $f\in X$, *for which the right-hand side of* (2.1.2) *tends to zero as* $n\to\infty$.

Further, if there exists M (an absolute constant) such that for n large enough,

$$\omega(1/n) \geqq M n^{-2\beta-2}, \tag{2.1.3}$$

then we have

THEOREM 2. *Let* $((\Delta\lambda_{n,k}))$ *be as in Theorem 1 and let* (2.1.3) *be satisfied. Then Theorem 1 holds with*

$$\|f(\cdot) - \sigma_n^{(\Lambda)}(f,(\cdot),X)\|_X \leqq$$

$$\leqq A\Big(\sum_{k=0}^{n} \frac{\omega(1/(k+1))}{(k+1)^{3/2+\alpha}} \sum_{\nu=0}^{k} (\nu+1)^{\alpha+1/2} \Delta\lambda_{n,\nu} L(\nu)\Big) \tag{2.1.4}$$

$(\alpha \geqq \beta \geqq -1/2)$, in place of (2.1.2).

REMARK 2. Condition (2.1.1) though very lighter in the sense that the ratio is allowed to be infinite with the order $n^{\alpha+1/2}$, is strict in the sense that it does not permit to use identity transformation (i.e. $\Delta\lambda_{n,k} = 0$ (for $n \neq k$) and $= 1$ (for $n = k$)), or a sequence with some terms zero. Changing our arguments a little we fill up the gap and find

THEOREM 3. *Let* $((\Delta\lambda_{n,k}))$, $(n = 0,1,\ldots,\ k = 0,1,2,\ldots)$ *be a lower triangular matrix with* $\sum_{\nu=0}^{n} \Delta\lambda_{n,\nu} = 1$ *and the sequence* $\{(k+1)^{\alpha+1/2}\Delta\lambda_{n,k}\}$ *be non-negative non-increasing for* $0 \leqq k \leqq n$. *Then* Λ*-transform processes of approximation of* f *through* (1.1.6) *have an order estimate*

$$\|f(\cdot) - \sigma_n^{(\Lambda)}(f,(\cdot),X)\|_X \leqq$$

$$\leqq A\Big(\sum_{k=0}^{n} \frac{\omega(1/(k+1))}{(k+1)^{3/2+\alpha}} \sum_{\nu=0}^{k} \Delta\lambda_{n,\nu}(\nu+1)^{\alpha+1/2} L(\nu) + n^{-2\beta-2}\Big) \tag{2.1.5}$$

or if (2.1.3) *is satisfied, then*

$$\leqq A\Big(\sum_{k=0}^{n} \frac{\omega(1/(k+1))}{(k+1)^{\alpha+3/2}} \sum_{\nu=0}^{k} \Delta\lambda_{n,\nu}(\nu+1)^{\alpha+1/2} L(\nu)\Big). \tag{2.1.6}$$

A is independent of f. The orders in (2.1.5), (2.1.6) *exhibit the property of saturation order if these orders tend to zero as $n \to \infty$. The respective saturation class or Favard's class $F(X,\sigma^{(\Lambda)})$ is the collection of those $f \in X$ for which the corresponding order tends to zero.*

REMARK 3. Since the non-increasingness of $\{(k+1)^{\alpha+1/2}\Delta\lambda_{n,k}\}$, $(\alpha \geqq -1/2,\ 0 \leqq k \leqq n)$ does not allow anyone to suppose that $\{\Delta\lambda_{n,k}\}$ is non-decreasing. Thus a conclusion similar to (1.3.6) is not possible in these processes.

It may be noted that a corollary similar to (1.3.11) does not hold for all α except $\alpha = -1/2$ which is not negligible at all. Thus we have by (2.1.2)

COROLLARY 2. *Let* $S_n^1(f, \cos\theta, X)$ *be* $(C,1)$*-transform of* (1.1.6). *Then for* $\alpha = \beta = -1/2$*:*

$$\|S_n^1(f,(\cdot),X) - f(\cdot)\|_X \leqq A\Big(\sum_{k=0}^{n} \frac{\omega(1/(k+1))}{(k+1)^{3/2+\alpha}} \sum_{\nu=0}^{k} \Delta\lambda_{n,\nu} + n^{-1}\Big) \leqq$$

(2.1.7)

$$\leqq An^{-1}\Big(\sum_{k=0}^{n} \omega(1/(k+1)) + O(1)\Big).$$

Corresponding result for (2.1.4) also holds by putting $\Delta\lambda_{n,\nu} = 1/(n+1)$ for $\nu \leqq n$ and zero otherwise, and leads Theorem C in the present case $\alpha = \beta = -1/2$ as well, i.e. if $\omega(1/n) \geqq Mn^{-1}$ then from (2.1.4) for $\alpha = \beta = -1/2$

$$\|f(\cdot) - S_n^1(f,(\cdot),X)\|_X \leqq An^{-1}\Big(\sum_{k=0}^{n} \omega(1/(k+1))\Big). \tag{2.1.8}$$

Thus for $\alpha = \beta = -1/2$ (Chebyshev polynomial of first kind) and any sequence $\{\Delta\lambda_{n,k}\}$ arising out of lower triangular matrix $((\Delta\lambda_{n,k}))$ with $\sum_{k=0}^{n} \Delta\lambda_{n,k} = 1$, we have

$$\|f(\cdot) - \sigma_n^{(\Lambda)}(f,(\cdot),X)\|_X \leqq A\Big(\sum_{k=0}^{n} |\Delta^2\lambda_{n,k}| \sum_{\nu=0}^{k} \omega(1/(\nu+1))\Big) \tag{2.1.9}$$

(by using (1.3.14)).

An appealing consequence of Theorems 1 and 2 is that now we can use the matrix (1.2.16) for $-1/2 \leqq \alpha \leqq 1/2$ which gives $(R, \log n, 1)$-mean of Jacobi series. Moreover, the case $\alpha \geqq \beta \geqq -1/2$, $-1/2 \leqq \alpha \leqq 1/2$ covers the important polynomials such as Chebyshev, Legendre, and ultraspherical.

Furthermore, the following concrete matrix leads a process of strong approximation for $f \in X$ through (1.1.6)

$$\Delta\lambda_{n,k} = \begin{cases} \dfrac{1}{(k+1)^{\mu} \sum_{j=0}^{n} \frac{1}{(j+1)^{\mu}}}, & k \leqq n, \\ 0, & k > n \end{cases} \tag{2.1.10}$$

when $\mu \geqq \alpha + 1/2$ and the saturation orders for this process are given by (2.1.2), (2.1.4), (2.1.5) and (2.1.6).

2.2 *Results to be used in the proof.* We shall need the following results to prove our theorems.

LEMMA 1. *Let* $\{(k+1)^{\alpha+1/2}\Delta\lambda_{n,k}\}$ *be non-negative non-increasing for* $0 \leqq k \leqq n$, *then for* $0 \leqq a < b \leqq \infty$, $0 < t < \pi$ *and for every* n,

$$\left|\sum_{a}^{b} \Delta\lambda_{n,k}(k+1)^{\alpha+1/2}e^{ikt}\right| \leqq A\sum_{k=0}^{\tau} \Delta\lambda_{n,k}(k+1)^{\alpha+1/2} \tag{2.2.1}$$

where τ *is the integral part of* $1/t$.

PROOF. If $\tau \leqq b$, we have

$$\left|\sum_{a}^{b} \Delta\lambda_{n,k}(k+1)^{\alpha+1/2}e^{ikt}\right| \leqq$$

$$\leqq \left|\sum_{a}^{\tau} \Delta\lambda_{n,k}(k+1)^{\alpha+1/2}e^{ikt}\right| + \left|\sum_{\tau}^{b} \Delta\lambda_{n,k}(k+1)^{\alpha+1/2}e^{ikt}\right| =$$

$$= I_1 + I_2 \quad \text{(say)},$$

where

$$I_1 \leqq \sum_{a}^{\tau} |\Delta\lambda_{n,k}(k+1)^{\alpha+1/2}| \quad (\text{for } |e^{ikt}| = 1)$$

$$\leqq \left|\sum_{k=0}^{\tau} \Delta\lambda_{n,k}(k+1)^{\alpha+1/2}\right| \quad \text{(for all terms are non-negative)},$$

and

$$I_2 \leqq \Delta\lambda_{n,\tau}\tau^{\alpha+1/2} \max_{\tau \leqq p < p' \leqq b} \left|\sum_{p}^{p'} e^{ikt}\right| \leqq A\tau\{\Delta\lambda_{n,\tau}\tau^{\alpha+1/2}\} \leqq$$

$$\leqq A\sum_{k=0}^{\tau} \Delta\lambda_{n,k}(k+1)^{\alpha+1/2}, \quad (\text{as } \{\Delta\lambda_{n,k}(k+1)^{\alpha+1/2}\} \searrow).$$

Thus (2.2.1) holds.

Again, if $\tau > b$

$$\left|\sum_{a}^{b} \Delta\lambda_{n,k}(k+1)^{\alpha+1/2}e^{ikt}\right| \leqq$$

$$\leqq \left|\sum_{a}^{\tau} \Delta\lambda_{n,k}(k+1)^{\alpha+1/2}e^{ikt}\right| + \left|\sum_{b}^{\tau} \Delta\lambda_{n,k}(k+1)^{\alpha+1/2}e^{ikt}\right| \leqq$$

$$\leqq A\sum_{a}^{\tau} \Delta\lambda_{n,k}(k+1)^{\alpha+1/2}$$

(as above in I_1). □

LEMMA 2. *If $\omega(\phi)$ is a non-negative real-valued function of $\phi \in [0, \pi]$ and for every $\varepsilon \geqq 0$, $\phi^{\varepsilon}\omega(\phi)$ is non-decreasing and tends to zero as $\phi \to 0$ then*

$$n^{-2\alpha-2}\omega(1/n)\sum_{\nu=0}^{n}\Delta\lambda_{n,\nu}(\nu+1)^{\alpha+1/2}L(\nu) \leqq$$

$$\leqq A\sum_{k=0}^{n}\frac{\omega(1/(k+1))}{(k+1)^{\alpha+3/2}}\sum_{\nu=0}^{k}(\nu+1)^{\alpha+1/2}\Delta\lambda_{n,\nu}L(\nu) \tag{2.2.2a}$$

for the non-negative sequence $\{\Delta\lambda_{n,k}\}$, $0 \leqq k \leqq n$. In particular, (2.2.2a) *holds for $\omega(\phi)$ defined by* (1.3.1).

If $\{(k+1)^{\alpha+1/2}\Delta\lambda_{n,k}\}$ is non-negative non-increasing with respect to k, $0 \leqq k \leqq n$, $\sum_{k=0}^{n}\Delta\lambda_{n,k} \leqq A \neq 0$, $\alpha \geqq \beta \geqq -1/2$, then

$$\omega(1/n) \leqq A\sum_{k=0}^{n}\frac{\omega(1/(k+1))}{(k+1)^{3/2+\alpha}}\sum_{\nu=0}^{k}(\nu+1)^{\alpha+1/2}\Delta\lambda_{n,\nu}L(\nu) \tag{2.2.2b}$$

for all positive integers n.

Further if (2.1.3) *is satisfied then*

$$n^{-2\beta-2} \leqq A\sum_{k=0}^{n}\frac{\omega(1/(k+1))}{(k+1)^{3/2+\alpha}}\sum_{\nu=0}^{k}(\nu+1)^{\alpha+1/2}\Delta\lambda_{n,\nu}L(\nu) \tag{2.2.3}$$

for n large enough.

PROOF. Proof of (2.2.2a) is a crystallized form of the proof of (3.9) in [8]. Also

$$\sum_{k=0}^{n}\frac{\omega(1/(k+1))}{(k+1)^{3/2+\alpha}}\sum_{\nu=0}^{k}(\nu+1)^{\alpha+1/2}\Delta\lambda_{n,\nu}L(\nu) \geqq$$

$$\geqq \sum_{k=0}^{n}\frac{\omega(1/(k+1))}{(k+1)^{3/2+\alpha}}(k+1)(k+1)^{\alpha+1/2}\Delta\lambda_{n,k},$$

$$(L(\nu) = 1 + O((\nu+1)^{-1}) = O(1)).$$

$$\geqq (1/A)\omega(1/n)\sum_{k=0}^{n}\Delta\lambda_{n,k} = A\omega(1/n)$$

(this proves (2.2.2b))

$$= A\omega(1/n)n^{2\beta+2}n^{-2\beta-2} \geqq n^{-2\beta-2} \qquad \text{(see (2.1.3)).}$$

This proves (2.2.3) in persuance of our convention that A is not the same at each occurrence.

Besides these lemmas many properties of Jacobi polynomials given in Szegő [7] are used. An important formula of Hilb's type is extracted as follows (see Bingham [4]).

For $C/n \leqq \theta \leqq \pi - C/n$, where n is large enough, $\alpha \geqq \beta \geqq -1/2$,

$$\omega_n^{(\alpha,\beta)} R_n^{(\alpha,\beta)}(\cos\theta) = \frac{2^{3/2}}{\pi^{1/2}\Gamma(\alpha+1)} n^{\alpha+1/2}(\sin\theta/2)^{-\alpha-1/2}\times \times(\cos\theta/2)^{-\beta-1/2}\cos\{n\theta+(\alpha+\beta+1)\theta/2\}L(n) \tag{2.2.4}$$

where $L(n) = 1 + O(1/n)$.

Proof of Theorem 1. From (1.2.10)

$$\|f(\cdot) - \sigma_n^{(\Lambda)}(f,(\cdot),X)\|_X \leqq \int_0^{\pi} \|T_\psi f(\cdot) - f(\cdot)\|_X |K_n^{(\Lambda)}(\psi)| P^{(\alpha,\beta)}(\psi)\, d\psi$$

$$= \int_0^{\pi/(n+1)} + \int_{\pi/(n+1)}^{\pi-\pi/(n+1)} + \int_{\pi-\pi/(n+1)}^{\pi} = P + Q + R \qquad \text{(say)}.$$

But

$$P \leqq \int_0^{\pi/(n+1)} \omega(\psi, f, X)|K_n^{(\Lambda)}(\psi)| P^{(\alpha,\beta)}(\psi)\, d\psi$$

$$\leqq A\omega(\pi/(n+1))(n+1)^{-2\alpha-2} \sum_{k=0}^{n} \Delta\lambda_{n,k} L_k \tag{2.2.5}$$

$$(L_k = k^{2\alpha+1}L(k)) \quad \text{see (1.2.8)}$$

$$\leqq A \sum_{k=0}^{n} \frac{\omega(1/(k+1))}{(k+1)^{3/2+\alpha}} \sum_{\nu=0}^{k} \Delta\lambda_{n,\nu}(\nu+1)^{\alpha+1/2} L(\nu)$$

(by Lemma 2, (2.2.2a) or (2.2.2b)).

Let n_0 be a number such that for $n \geqq n_0$, the order given by (2.2.4) holds,

then

$$Q \leqq \int_{\pi/(n+1)}^{\pi-\pi/(n+1)} \omega(\psi, f, X)\Big|\sum_{k=0}^{n_0} \Delta\lambda_{n,k} L_k R_k^{(\alpha+1,\beta)}(\cos\psi)\Big| P^{(\alpha,\beta)}(\psi)\, d\psi$$
$$+ \int_{\pi/(n+1)}^{\pi-\pi/(n+1)} \omega(\psi, f, X)\Big|\sum_{k=n_0+1}^{n} \Delta\lambda_{n,k} L_k R_k^{(\alpha+1,\beta)}(\cos\psi)\Big| P^{(\alpha,\beta)}(\psi)\, d\psi$$
$$= Q_1 + Q_2 \qquad \text{(say)}$$

such that

$$Q_1 \leqq A \int_{\pi/(n+1)}^{\pi} \psi^{2\beta+1}\omega(\psi, f, X) \sum_{k=0}^{n_0} \Delta\lambda_{n,k}\, d\psi$$
$$\leqq A n_0 \Delta\lambda_{n,0} \sum_{k=0}^{n} \int_{\pi/(k+2)}^{\pi/(k+1)} \psi^{2\alpha+1}\omega(\psi)\, d\psi \frac{(k+1)\Delta\lambda_{n,k}(k+1)^{\alpha+1/2}L(k)}{(k+1)\Delta\lambda_{n,k}(k+1)^{\alpha+1/2}L(k)}\, d\psi$$

for $\Delta\lambda_{n,k} \neq 0$, $k = 0, 1, 2, \ldots, n$) (see (2.1.1)).

$$\leqq A\Delta\lambda_{n,0} \sum_{k=0}^{n} \frac{\omega(1/(k+1))}{(k+1)^{3/2+\alpha}} \frac{(k+1)^{-3/2-\alpha}}{(k+1)^{\alpha+3/2}\Delta\lambda_{n,k}} \sum_{\nu=0}^{k} \Delta\lambda_{n,\nu}(\nu+1)^{\alpha+1/2}L(\nu)$$
$$\leqq A\Big[\frac{\Delta\lambda_{n,0}}{\Delta\lambda_{n,n}}\Big] n^{-\alpha-1/2} \sum_{k=0}^{n} \frac{\omega(1/(k+1))}{(k+1)^{3/2+\alpha}} (k+1)^{-5/2-\alpha} \times$$
$$\times \sum_{\nu=0}^{k} \Delta\lambda_{n,\nu}(\nu+1)^{\alpha+1/2}L(\nu)$$

(for $\{(k+1)^{\alpha+1/2}\Delta\lambda_{n,k}\} \searrow \Rightarrow 1/\{(k+1)^{\alpha+1/2}\Delta\lambda_{n,k}\} \nearrow$)

$$\leqq A \sum_{k=0}^{n} \frac{\omega(1/(k+1))}{(k+1)^{3/2+\alpha}} \sum_{\nu=0}^{k} \Delta\lambda_{n,\nu}(\nu+1)^{\alpha+1/2}L(\nu) \qquad \text{(by (2.1.1))}. \tag{2.2.6}$$

Again for $\alpha \geqq \beta \geqq -1/2$

$$Q_2 = \frac{2^{3/2}}{\pi^{1/2}\Gamma(\alpha+1)} \times$$

$$\times \int_{\pi/(n+1)}^{\pi-\pi/(n+1)} \omega(\psi) \Big| \sum_{k=n_0+1}^{n} \Delta\lambda_{n,k} \frac{k^{\alpha+3/2} L(k)}{2k+\alpha+\beta+2} \cos\Big(k\psi + \frac{\alpha+\beta+2}{2}\psi\Big)\Big| \times$$

$$\times \Big(\sin\frac{\psi}{2}\Big)^{\alpha-1/2} \Big(\cos\frac{\psi}{2}\Big)^{\beta+1/2} d\psi \leqq$$

$$\leqq A \int_{\pi/(n+1)}^{\pi} \frac{\omega(\psi)}{\psi^{1/2-\alpha}} \sum_{\nu=0}^{[1/\psi]} \Delta\lambda_{n,\nu} (\nu+1)^{\alpha+1/2} L(\nu) \qquad \text{(by (2.2.1))}$$

(for the integrand is positive)

$$\leqq A \sum_{k=0}^{n} \int_{\pi/(k+2)}^{\pi/(k+1)} \frac{\omega(\psi)}{\psi^{1/2-\alpha}} \sum_{\nu=0}^{[1/\psi]} \Delta\lambda_{n,\nu} (\nu+1)^{\alpha+1/2} L(\nu) \leqq$$

(2.2.7)

$$\leqq A \sum_{k=0}^{n} \frac{\omega(1/(k+1))}{(k+1)^{3/2+\alpha}} \sum_{\nu=0}^{k} (\nu+1)^{\alpha+1/2} \Delta\lambda_{n,\nu} L(\nu).$$

Again

$$R \leqq \int_{\pi-\pi/(n+1)}^{\pi} \omega(\psi) |K_n^{(\Lambda)}(\psi)| P^{(\alpha,\beta)}(\psi)\, d\psi.$$

But

$$K_n^{(\Lambda)}(\psi) \equiv F(\cos\psi) = \sum_{k=0}^{n} \Delta\lambda_{n,k} L_k R_k^{(\alpha+1,\beta)}(\cos\psi) =$$

$$= \sum_{k=1}^{n} b(k) k^{-1} \omega_k^{(\alpha+1,\beta)} R_k^{(\alpha+1,\beta)}(\cos\psi) + b(0),$$

where

$$b(k) = \begin{cases} \dfrac{(\alpha+1)\Delta\lambda_{n,k}}{2+\frac{\alpha+\beta+2}{k}}, & k > 0; \\ \dfrac{(\alpha+1)\Delta\lambda_{n,0}}{\alpha+\beta+2}, & k = 0. \end{cases}$$

Thus with an application of an argument ([3], pp. 785–86), we see that $F(\cos\psi)$ is continuous in $0 < \psi \leqq \pi$ (e.g. see the abstract of [3]). Thus for n

large enough

$$
\begin{aligned}
R &\leqq A \int_{\pi-\pi/(n+1)}^{\pi} \omega(\psi)\psi^{2\alpha+1}(\pi-\psi)^{2\beta+1}\,d\psi = \\
&= A \int_{0}^{\pi/(n+1)} \psi^{2\beta+1}\,d\psi = An^{-2\beta-2}.
\end{aligned} \tag{2.2.8}
$$

Further argument similar to (1.3.10) holds and we conclude that (2.1.2) is saturation order, because

$$
\begin{aligned}
&\int_{0}^{\pi} |K_n^{(\Lambda)}(\psi)| P^{(\alpha,\beta)}(\psi)\,d\psi \geqq An^{\alpha+3/2}\Delta\lambda_{n,0} \geqq \qquad \text{(for some } A>0) \\
&\geqq A\sum_{k=0}^{n} \frac{\omega(1/(k+1))}{(k+1)^{\alpha+3/2}} \sum_{\nu=0}^{k} \Delta\lambda_{n,\nu}(\nu+1)^{\alpha+1/2}L(\nu)
\end{aligned} \tag{2.2.9}
$$

for some non-trivial $f \in X$. This proves Theorem 1.

REMARK 4. It may be remarked that our all these estimates are the orders of some strong approximation processes only when they tends to zero as $n \to \infty$. And these orders become saturation orders if $\Delta\lambda_{n,k}$ are (non-zero) positive for all n, k and the corresponding conditions are satisfied.

PROOF OF THEOREM 2. The proof of this theorem follows on the lines of the proof of Theorem 1. But in the present case we use (2.2.3) of Lemma 2 to get a single term in the right-hand side of (2.1.4).

PROOF OF THEOREM 3. If we do not use the condition (2.1.1) which is only used in the proof of Q_1, we remember that

$$
\begin{aligned}
Q_1 &\leqq A\sum_{k=0}^{n_0} \Delta\lambda_{n,k} \leqq A\sum_{k=0}^{n_0} (k+1)^{\alpha+1/2}\Delta\lambda_{n,k}L(k) \leqq \\
&\leqq A\sum_{k=0}^{n_0} \frac{\omega(1/(k+1))}{(k+1)^{\alpha+3/2}} \sum_{\nu=0}^{k} (\nu+1)^{\alpha+1/2}\Delta\lambda_{n,\nu}L(\nu) \leqq \\
&\leqq A\sum_{k=0}^{n} \frac{\omega(1/(k+1))}{(k+1)^{\alpha+3/2}} \sum_{\nu=0}^{k} (\nu+1)^{\alpha+1/2}\Delta\lambda_{n,\nu}L(\nu) \qquad (n > n_0)
\end{aligned} \tag{2.2.10}
$$

where A depends upon f. With this the proofs of Theorem 1, 2 furnishes the proofs of (2.1.5) and (2.1.6). □

REMARK 5. From Theorems 1, 2 and 3 and the theorems of [8] we conclude that the monotonicity of $\{(k+1)^{\alpha+1/2}\Delta\lambda_{n,k}\}$ $(\alpha \geqq \beta \geqq -1/2)$ generates the processes of strong approximation in X and the orders obtained are saturation orders.

REFERENCES

[1] ASKEY, R. and WAINGER, S., A convolution structure for Jacobi series, *Amer. J. Math.* **91** (1969), 463–485. *MR* **41**#8728

[2] BAVINCK, H., Approximation processes for Fourier–Jacobi expansions, *Applicable Anal.* **5** (1976), 293–312. *MR* **54**#3243

[3] BAVINCK, H., A special class of Jacobi series and some applications, *J. Math. Anal. Appl.* **37** (1972), 767–797. *MR* **46**#7589

[4] BINGHAM, N. H., Tauberian theorems for Jacobi series, *Proc. London Math. Soc.* (3) **36** (1978), 285–309. *MR* **58**#29795

[5] BUTZER P. L. and BERENS, H., *Semi-groups of operators and approximation*, Die Grundlehren der mathematischen Wissenschaften, Bd. 145, Springer-Verlag, New York, 1967. *MR* **37**#5588

[6] GASPER, G., Positivity and the convolution structure for Jacobi series, *Ann. of Math.* (2) **93** (1971), 112–118. *MR* **44**#1852

[7] SZEGŐ, G., *Orthogonal polynomials*, Third edition, American Mathematical Society Colloquium Publications, Vol. 23, American Mathematical Society, Providence, R. I., 1967. *MR* **46**#9631

[8] YADAV, S. P., On the saturation order of approximation processes involving Jacobi polynomials, *J. Approx. Theory* **58** (1989), 36–49. *MR* **91c**:41065

(Received February 5, 1990)

DEPARTMENT OF MATHEMATICS
GOVERNMENT MAHARAJA COLLEGE
A.P.S. UNIVERSITY, REWA
IND–471 001 CHHATARPUR, M.P.
INDIA

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 317–320

ON THE ACTION OF p'-AUTOMORPHISMS ON p-GROUPS HAVING SOFT SUBGROUPS

L. HÉTHELYI

In this paper we shall derive some properties of the action of p'-automorphisms on p-groups with soft subgroups. The notion of soft subgroup was introduced in [1] where the basic properties of p-groups having soft subgroups were established. In this paper we show that the action of a p'-group on a p-group with soft subgroups is very similar to that of the action of a p'-group acting on a p-group generated by two elements. We establish some general properties of the action of p'-automorphism on p-groups and then investigate them in the case of p-groups with soft subgroups.

PROPOSITION 1. *Suppose G is a p-group, $A \leqq \mathrm{Aut}(G)$, $(p, |A|) = 1$ and $G' \leqq C_G(A)$. Then A acts trivially on $G/Z_2(G)$.*

PROOF. Let $C_2 = C_G(A)$, $C = C_G(G')$. Then $C_2 \triangleleft G$ and $\mathrm{cl}(C) \leqq 2$ for $C' \leqq Z(C)$. As $[G, G', A] = [G', A, G] = 1$ we have $[A, G, G'] = 1$ by the Three Subgroup Lemma. Thus $[G, A] \leqq C$ and $[G, A] = [C, A]$. Let $C_1 = [C, A]$. Then $C_1 \triangleleft G$ and $G = C_1 C_2$. As $[C_1, C_2, A] = [C_2, A, C_1] = 1$ we have $[C_1, C_2] = [A, C_1, C_2] = 1$ by the Three Subgroup Lemma. We now show that $C_1 \leqq Z_2(G)$.

It is easy to see that $Z(C_1) \leqq Z(G)$. Let a bar denote homomorphic images in $G/Z(C_1)$. Then $\bar{G} = \bar{C}_1 \cdot \bar{C}_2$, $[\bar{C}_1, \bar{C}_2] = 1$ and so as $\bar{C}_1$ is abelian $\bar{C}_1 \leqq Z(\bar{G})$. Thus $C_1 \leqq Z_2(G)$ which proves the proposition. Q.E.D.

COROLLARY 1. *Suppose G is a p-group, $A \leqq \mathrm{Aut}(G)$, $(p, |A|) = 1$. If A acts trivially on $G' \cdot Z_2(G)$ then A is trivial on G.*

COROLLARY 2. *Suppose G is a p-group, $\mathrm{cl}(G) \geqq 3$, $A \leqq \mathrm{Aut}(G)$, $(p, |A|) = 1$. If A acts trivially on $Z_n(G)$ (where $Z_n(G)$ is the last proper term of the upper central series of G) then A is trivial on G.*

COROLLARY 3. *Suppose G is a p-group generated by two elements, and that $\mathrm{cl}(G) \geqq 3$. Suppose $A \leqq \mathrm{Aut}(G)$, $(p, |A|) = 1$. If A acts trivially on G' then A is trivial on G.*

PROOF. As $\mathrm{cl}(G) \geqq 3$, $Z_2(G) < \Phi(G)$. Then A is trivial on $G/\Phi(G)$, so A is trivial on G. Q.E.D.

We now prove a partial converse of Proposition 1.

1991 *Mathematics Subject Classifications.* Primary 20D15; Secondary 20D45.
Key words and phrases. Group, automorphism, maximal abelian group, commutator.

Akadémiai Kiadó, Budapest

PROPOSITION 2. *Suppose G is a p-group, $A \leqq \mathrm{Aut}(G)$, $(p, |A|) = 1$. If A acts trivially on $G/Z_2(G)$ then A acts trivially on $K_3(G)$.*

PROOF. Let the bar denote homomorphic images in $G/Z(G)$. Then $[\bar{G}, A] < Z(\bar{G})$ and thus $[\bar{G}, A] = 1$ so $[G', A] \leqq Z(G)$ and $[G', A, G] = 1$ follows. Moreover, $[A, G, G'] = 1$. Thus we have $[G, G', A] = 1$ by the Three Subgroup Lemma. Q.E.D.

We shall now investigate some of the action of p'-automorphisms of a p-group with soft subgroups.

LEMMA 1. *Suppose G is a p-group, B is a soft subgroup of G of index at least p^2. Let $R(G) = G' \cdot Z(N_G(B))$. Then $Z_2(G)$ is abelian and $R(G) \leqq C_G(Z_2(G))$.*

PROOF. If B_1 is any soft subgroup of G of index at least p^2 then $Z_2(G) \leqq N_G(B_1)$. In particular, if M is the unique maximal subgroup of G containing B then $Z_2(G) \leqq Z_2(M)$ by [1]. Moreover, $|Z_2(G) \cdot Z(M) : Z(M)| \leqq p$ which shows that $Z_2(G) \leqq R(G) \cap N_G(B)$. As $Z_2(G) \leqq C_G(G'))$ and $Z_2(G) \leqq C_G(Z(N_G(B)))$ we have $Z_2(G) \leqq Z(R(G))$. Q.E.D.

PROPOSITION 3. *Suppose G is a p-group, B is a soft subgroup of G of index at least p^2. Suppose $A \leqq \mathrm{Aut}(G)$ acting trivially on $R(G) = G' \cdot Z(N_G(B))$. Then A is trivial on G.*

PROOF. The proposition follows from Proposition 1 and Lemma 1. Q.E.D.

PROPOSITION 4. *Suppose G is a p-group, B is a soft subgroup of G of index at least p^2. Let $R(G) = G' \cdot Z(N_G(B))$. Then either $C_G(R(G)) \leqq R(G)$ or $C_G(R(G))$ is of index p in G containing $R(G)$.*

PROOF. Suppose $C_G(R(G)) \not\leqq R(G)$. Let M be the unique maximal subgroup of G containing B. If $C_G(R(G)) \leqq M$ then there is an element g in $B \setminus R(G)$ such that $g \in C_G(R(G))$. So $R(G) \leqq B$ and then $R(G) \leqq Z(M)$. Thus $M' = 1$ which does not hold. So $C_G(R(G)) \not\leqq M$.

In particular G' is abelian. Then $G' \cdot Z(M)$ is a maximal abelian subgroup of M contained in $R(G)$ by Statement 3 of [2]. However, both G' and $Z(M)$ centralize $R(G)$ which means that $G' \cdot Z(M) = R(G)$. Q.E.D.

PROPOSITION 5. *Suppose G is a p-group and B is a soft subgroup of G of index at least p^2. Suppose $A \leqq \mathrm{Aut}(G)$ and $(p, |A|) = 1$ and that A acts trivially on G'. Then $[G, A] \leqq Z(G)$.*

PROOF. Let $C = C_G(G')$, and M be the unique maximal subgroup containing B. By the Three Subgroup Lemma we have $[G, A] < C$. Let $C_1 = [G, A]$, $C_2 = C_G(A)$. If $C_1' = 1$ then as $[C_1, C_2] = 1$, $C_1 \leqq Z(G)$ would follow. Thus we can suppose that $C_1' \neq 1$ and so $C' \neq 1$ and so $C \not\leqq M$ by Statement 3 of [2]. So $C \cdot B = G$ and so $G' \leqq C$, which means that G' is abelian and $C \cap M = G' \cdot Z(M)$. Then $C = G' \cdot Z(M)\langle y \rangle$ for some $y \in C \setminus M$.

Thus $C/Z(M)\cdot G'$ is cyclic and as $G' \leqq Z(C)$, $C/Z(M)$ is abelian so $C' \leqq$ $\leqq Z(M)$. As $C_1' \neq 1$ but $C \cap M$ is abelian, $C_1 \not\leqq M$ follows. Then $C_1' \leqq G'$. If $G' \neq C_1'$ then $[C_1/C_1', A] < C_1/C_1'$ and thus $[C_1, A] < C_1$ which is not the case. So $G' = C_1' \leqq Z(M)$. So $B \triangleleft G$ which is a contradiction. Q.E.D.

COROLLARY 5. *Suppose G is a p-group, B is a soft subgroup of G of index at least p^3. Let M be the unique maximal subgroup of G containing B. Let the bar denote homomorphic images in $G/Z(M)$. Let $A \leqq \mathrm{Aut}(\bar{G})$, $(p, |A|) = 1$.*

If A acts trivially on $\bar{G}'$, then A acts trivially on $\bar{G}$.

PROOF. By Proposition 5 A is trivial on $\bar{G}/Z(\bar{G})$. However, as $|Z(\bar{G})| =$ $= p$, $Z(\bar{G}) \leqq \bar{G}' \leqq \Phi(\bar{G})$. Q.E.D.

PROPOSITION 6. *Suppose G is a p-group, B is a soft subgroup of G of index at least p^2. Let M be the unique maximal subgroup of G containing B. Let $R(G) = G' \cdot Z(N_G(B))$. Suppose $A \leqq \mathrm{Aut}(G)$, $(p, |A|) = 1$ and that A is trivial on $G/R(G)$. Then $[G, A] \leqq Z(M)$.*

PROOF. We first prove that A centralizes M'. As $[G, A] \leqq R(G)$, $R(G)\cdot$ $\cdot C_G(A) \geqq [G, A]\cdot C_G(A) = G$. Thus $C_G(A) \cap M \not\leqq R(G)$. So by Propositions 3 and 4 in [2] there exists an element a of M in $C_G(A)$ such that $\langle a, a^b\rangle' = M'$ for any $b \in C_G(A) \setminus M$. Now $[R(G), A] = [G, A]$ so $[G, A] = [M, A]$. Thus by Proposition 5 $[G, A] \leqq Z(M)$. Q.E.D.

COROLLARY 6. *Suppose G is a p-group, B is a soft subgroup of B of index at least p^3. Let M be the unique maximal subgroup of G containing B. Let a bar denote homomorphic images in $G/Z(M)$. Let $A \leqq \mathrm{Aut}(\bar{G})$, $(p, |A|) = 1$.*

If A acts trivially on $\bar{G}/R(\bar{G})$ then A acts trivially on $\bar{G}$.

PROOF. A is trivial on $\bar{G}/Z(\bar{M})$ by Proposition 6. So A is trivial on $\bar{M}'$. As $|Z(\bar{M})| \leqq p^2$ and as $|\bar{M}' \cap Z(\bar{M})| \geqq p$, $|[Z(\bar{M}), A]| \leqq p$. However, $[Z(\bar{M}), A] \triangleleft \bar{G}$ so $[Z(\bar{M}), A] \leqq Z(\bar{G}) \leqq \bar{G}' \leqq \Phi(\bar{G})$. Thus A is trivial on $\bar{G}$. Q.E.D.

PROPOSITION 7. *Suppose G is a p-group and B is a soft subgroup of G of index at least p^2. Let M be the unique maximal subgroup of G containing B. Let $R(G) = G' \cdot Z(N_G(B))$. Let N be an abelian normal subgroup of G not contained in M. Then $R(G)\cdot N$ is a characteristic subgroup of G of index p.*

PROOF. Let $L = Z_2(G) \cap G'$. Then $L < N$ and both $R(G)$ and N centralizes L. As $Z_2(G) \cap R(G) > Z(G) \cap G'$, $R(G)\cdot N$ is a proper subgroup of G. Moreover, $N \cap M = N \cap R(G)$ and thus $R(G)\cdot N$ is of index p in G. Thus $R(G)\cdot N$ is a characteristic subgroup of G of index p. Q.E.D.

Finally we shall examine the actions of p'-automorphisms on a p-group of maximal class.

LEMMA 2. *Suppose G is a non-exceptional p-group of maximal class. Suppose $A \leqq \mathrm{Aut}(G)$, $(p, |A|) = 1$ and that A is trivial on $Z_2(G)$. Then A is trivial on G.*

PROOF. Let $C = C_G(Z_2(G))$ and $C_1 = [G, A]$. As $[A, Z_2(G), G] = [Z_2(G), G, A] = 1$, $C_1 \leqq C$ by the Three Subgroup Lemma. Then $C_1 = [C, A]$. Let $C_2 = C_G(A)$. Then $C_1 \cdot C_2 = G$ and so we can assume that $C_1 \not\leqq G'$. Thus $|G : C_1| = p$ and $C_2 \cdot G' \leqq G$ is of maximal class. Also $Z(C_2 \cdot G') = Z(G)$. Thus $\langle c \rangle \times A$ acts on G' and the centralizer of $\langle c \rangle$ in G' is centralized by A where c is an arbitrary element of $C_2 \setminus G'$. Then by the Thompson Direct Product Lemma, A acts trivially on G' and then on G by Corollary 3. Q.E.D.

PROPOSITION 8. *Suppose G is a p-group of maximal class of order at least p^4. Let $A \leqq \mathrm{Aut}(G)$, $(p, |A|) = 1$. If A acts trivially on $Z_3(G)$ then A is trivial on G.*

PROOF. Let $G = G/Z(G)$. Then G is a non-exceptional p-group of maximal class. So the Proposition follows from Lemma 2. Q.E.D.

REFERENCES

[1] HÉTHELYI, L., Soft subgroups of p-groups, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **27** (1984), 81–85. *MR* **87c**: 20044

[2] HÉTHELYI, L., On subgroups of p-groups having soft subgroups, *J. London Math. Soc.* (2) **41** (1990), 425–437.

[3] HUPPERT, B., *Endliche Gruppen.* I, Die Grundlehren der mathematischen Wissenschaften, Bd. 134, Springer-Verlag, Berlin–New York, 1967. *MR* **37** #302

(Received February 8, 1990)

BUDAPESTI MŰSZAKI EGYETEM
VEGYÉSZMÉRNÖKI KAR
MATEMATIKA TANSZÉK
EGRI JÓZSEF U. 1
H–1521 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 321–338

ON THE ESTIMATE OF $(x_{\min} + x_{\max})/2$. II

I. JOÓ and S. SZABÓ

In [1] the estimate $(x_{\min} + x_{\max})/2$ has been investigated for symmetrical distributions. The aim of the present note is to extend these results for non-symmetrical case.

The following problem often occurs in the practice: we have to measure some quantity and the results of n independent measurements are $x_1, \ldots, x_n$ (real numbers). Give a "good" approximation for the quantity considered, using the values x_i. Usually we consider $n^{-1}(x_1 + \ldots + x_n)$ as an approximation. Another possible approximation is, e.g. $2^{-1}(x_{\min} + x_{\max})$, where $x_{\min} := \min(x_1, \ldots, x_n), x_{\max} := \max(x_1, \ldots, x_n)$. This approximation is very sensitive with respect to the errors, hence it is improbable that for $n \to \infty$ the exactness of the estimate increases. We have investigated this estimate in [1] for the case that $x_1, \ldots, x_n$ are the values of a symmetrical random variable and we have found that in some cases (e.g. for $F(x) = x + \frac{1}{2}$, $-\frac{1}{2} \leqq x \leqq \frac{1}{2}$, which is the special case of V in [1], when $\beta = \gamma = c_9 = 1, \delta = 0, h = 1/2$) this estimate is better than that of the arithmetic mean $n^{-1}(x_1 + \ldots + x_n)$. Here we investigate some cases, when $x_1, \ldots, x_n$ are the values of a non-symmetrical random variable. We consider only some special non-symmetrical distribution. First we prove the following

THEOREM A. *Let F be a distribution function on R such that $F(x) = 0$ if $x \leqq 0$, and $0 < F(x) < 1$ if $x > 0$. Suppose F is absolute continuous and denote $f(x) = F'(x)$. Fix $0 < b < 1$. Then for the solution y of the equation*

$$n \int_{-\infty}^{y} [F(2y - x) - F(x)]^{n-1} f(x)dx = b$$

we have

$$F(2y) = \sqrt[n]{b} + O(D^n) + O\Big(\max_{2y-1 \leqq t \leqq 2y} f(t)\Big)$$

where $0 < D < 1$ is a fixed constant, independent of b and of n.

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 62F35; Secondary 62H10.

Key words and phrases. Distribution function, density function.

Akadémiai Kiadó, Budapest

PROOF. (a) $y>0$ because $f(x)=0$ if $x\leqq 0$. First we prove that $y\to\infty$. Indeed, if $0\leqq y_1<y_2$ then

$$n\int_0^{y_1}[F(2y_1-x)-F(x)]^{n-1}f(x)dx\leqq n\int_0^{y_2}[F(2y_2-x)-F(x)]^{n-1}f(x)dx.$$

If we suppose indirectly that $y=O(1)$ then we obtain

$$b=n\int_0^{y}[F(2y-x)-F(x)]^{n-1}f(x)dx\leqq n\int_0^{O(1)}[F(O(1)-x)-F(x)]^{n-1}f(x)dx$$

$$\leqq n\int_0^{O(1)}d^{n-1}f(x)dx=O(nd^{n-1}),$$

where $0<d<1$ is a fixed constant. But $b=O(nd^{n-1})$ is a contradiction.

(b) We shall show that

$$n\int_0^{y}[F(2y-x)-F(x)]^{n-1}f(x)dx=$$

$$=n\int_0^{1}\left[F(2y)-c_1\max_{2y-1\leqq t\leqq 2y}f(t)-F(x)\right]^{n-1}f(x)dx+c_2n(1-F(1))^{n-1},$$

$0\leqq c_1,\ c_2\leqq 1$, where instead of 1 in the upper bound of the last integral we can write arbitrary positive real number. According to $y\to\infty$ $(n\to\infty)$ we can assume $y>1$, namely this holds for $n\geqq n_0$. Obviously, we have for $y>1$

$$n\int_0^{y}[F(2y-x)-F(x)]^{n-1}f(x)dx=$$

$$=n\int_0^{1}[F(2y-x)-F(x)]^{n-1}f(x)dx+n\int_1^{y}[F(2y-x)-F(x)]^{n-1}f(x)dx=I_1+I_2,$$

further

$$I_2=n\int_1^{y}[F(2y-x)-F(x)]^{n-1}f(x)dx\leqq n\int_1^{y}[1-F(1)]^{n-1}f(x)dx\leqq n(1-F(1))^n.$$

According to the Lagrange inequality

$$F(2y)-F(2y-x)\leqq c_1 \max_{2y-1\leqq t\leqq 2y} f(t), \qquad (0\leqq x\leqq 1)$$

and summarizing our estimates we have

$$n\int_0^y [F(2y-x)-F(x)]^{n-1} f(x)dx =$$

(1)

$$=n\int_0^1 [F(2y)-c_1 \max_{2y-1\leqq t\leqq 2y} f(t)-F(x)]^{n-1} f(x)dx+c_2 n(1-F(1))^{n-1}.$$

(c) Now we give the desired estimate for y. From (1) we get

$$b=n\int_0^y [F(2y-x)-F(x)]^{n-1} f(x)dx=[F(2y)-c_1 \max_{2y-1\leqq t\leqq 2y} f(t)]^n-$$

$$-[F(2y)-F(1)-c_1 \max_{2y-1\leqq t\leqq 2y} f(t)]^n+c_2 n(1-F(1))^{n-1}.$$

Here

$$0<F(2y)-F(1)-c_1 \max_{2y-1\leqq t\leqq 2y} f(t)\leqq B<1 \qquad (n\geqq n_1),$$

where B is a fixed constant and n_1 is an effective constant. So we have

$$F(2y)=\sqrt[n]{b}(1+O((1-F(1))^{n-1})+O(n^{-1}B^n))+O(\max_{2y-1\leqq t\leqq 2y} f(t))$$

and Theorem A is proved.

REMARK 1. In the application we have to estimate y before we apply Theorem A, using the ideas of its proof. Namely, we saw that

$$n\int_0^y [F(2y-x)-F(x)]^{n-1} f(x)dx = n\int_0^1 [F(2y-x)-F(x)]^{n-1} f(x)dx+$$

$$+c_2(n(1-F(1))^{n-1}),$$

where

$$n\int_0^1 [F(2y-x)-F(x)]^{n-1} f(x)dx \leqq n\int_0^1 [F(2y)-F(x)]^{n-1} f(x)dx =$$

$$= (F(2y))^n - (F(2y) - F(1))^n,$$

hence

$$b = n \int_0^y [F(2y - x) - F(x)]^{n-1} f(x)dx \leqq (F(2y))^n + n(1 - F(1))^n,$$

consequently

$$F(2y) \geqq \sqrt[n]{b} + O((1 - F(1))^n), \quad n \geqq 1. \tag{2}$$

REMARK 2. Obviously, $\sqrt[n]{b} = 1 + \frac{\log b}{n} + O(n^{-2})$ so in the case

$$\max_{2y-1 \leqq t \leqq 2y} f(t) \geqq c/n \quad \text{(for some } c > 0)$$

we have to modify the calculations in order to obtain better estimate. Let $0 < \varepsilon_n$, $\varepsilon_n \to 0$ $(n \to \infty)$, (we will choose it later) and consider the partition

$$n \int_0^y [F(2y - x) - F(x)]^{n-1} f(x)dx = n \int_0^{\varepsilon_n} [F(2y - x) - F(x)]^{n-1} f(x)dx +$$

$$+ n \int_{\varepsilon_n}^y [F(2y - x) - F(x)]^{n-1} f(x)dx = I_1 + I_2.$$

Obviously,

$$I_2 = n \int_{\varepsilon_n}^y [F(2y - x) - F(x)]^{n-1} f(x)dx \leqq (1 - F(\varepsilon_n))^n,$$

$$F(2y) - F(2y - x) = c_1 \varepsilon_n \max_{2y-\varepsilon_n \leqq t \leqq 2y} f(t), \quad 0 \leqq x \leqq \varepsilon_n, \ 0 \leqq c_1 \leqq 1,$$

consequently,

$$n \int_0^y [F(2y - x) - F(x)]^{n-1} f(x)dx =$$

$$= n \int_0^{\varepsilon_n} [F(2y) - c_1 \varepsilon_n \max_{2y-\varepsilon_n \leqq t \leqq 2y} f(t) - F(x)]^{n-1} f(x)dx + c_3(1 - F(\varepsilon_n))^n, \tag{3}$$

and hence

$$b = n\int_0^y [F(2y-x)-F(x)]^{n-1} f(x)dx = [F(2y)-c_1\varepsilon_n \max_{2y-\varepsilon_n \leqq t \leqq 2y} f(t)]^n -$$

$$-c_4\Big([1-F(\varepsilon_n)-c_1\varepsilon_n \max_{2y-\varepsilon_n \leqq t \leqq 2y} f(t)]^n\Big) + c_3(1-F(\varepsilon_n))^n.$$

Now choose $\varepsilon_n \to 0$ so that if possible $(1-F(\varepsilon_n))^n \to 0$, further

$$\varepsilon_n \max_{2y-\varepsilon_n \leqq t \leqq 2y} f(t) = o(n^{-1})$$

be fullfilled. We obtain

THEOREM B. *Under the assumptions of Theorem A we have*

$$F(2y) = \sqrt[n]{b} + O(\varepsilon_n \max_{2y-\varepsilon_n \leqq t \leqq 2y} f(t)) + O(n^{-1}(1-F(\varepsilon_n))^n).$$

The following theorem is true for probability distribution functions with finite and infinite support of density functions.

THEOREM C. *Suppose F is a probability distribution function such that $F(x)=0$ for $x \leqq 0$, $F>0$ for $x>0$ and F is continuous at 0. Suppose $0<\alpha(n)$, $\beta(n)$ are monotone increasing sequences, tending to $+\infty$ arbitrarily slowly and suppose $\alpha(n)=o(n)$. Let ε_n and δ_n be such that $F(\varepsilon_n)=\alpha(n)/n$, $F(\delta_n)=\frac{1}{n\beta(n)}$. Then*

$$\sqrt[n]{b}\Big(1+O\Big(\frac{e^{-\alpha(n)}}{n}\Big)+O\Big(\frac{1}{n\beta(n)}\Big)\Big) \geqq F(2y-\varepsilon_n), \tag{4}$$

$$\sqrt[n]{b}\Big(1+O\Big(\frac{e^{-\alpha(n)}}{n}\Big)+O\Big(\frac{1}{n\beta(n)}\Big)\Big) \geqq F(2y-\delta_n). \tag{5}$$

PROOF. Denote F the smallest closed interval containing supp f. We may suppose without loss of generality that in the case that supp f is finite $F=[0,1]$.

(i) The case $F=[0,1]$.

(a) $y>0$, because $f(x)=0$ if $x<0$. First we prove that $y \leqq q < 1/2$ is false, where $q>1/3$ is an absolute constant. Indeed, if $0 \leqq y_1 < y_2 \leqq 1/2$ then

$$n\int_0^{y_1} [F(2y_1-x)-F(x)]^{n-1} f(x)dx \leqq n\int_0^{y_2} [F(2y_2-x)-F(x)]^{n-1} f(x)dx.$$

If we suppose that $y \leqq q < 1/2$ (q is an absolute constant, $q > 1/3$), then

$$b = n \int_0^y [F(2y-x) - F(x)]^{n-1} f(x)dx \leqq$$

$$\leqq n \int_0^{1/2} [F(2q-x) - F(x)]^{n-1} f(x)dx \leqq$$

$$\leqq n \int_0^{1/2} [F(2q) - F(x)]^{n-1} f(x)dx \leqq 2F^n(2q).$$

But $b \leqq 2F^n(2q)$ is a contradiction.

(b) Let ε_n, δ_n be such that $0 < \delta_n < \varepsilon_n < 1/4$, $\varepsilon_n \to 0$. We will choose ε_n, δ_n below.

$$b = n \int_0^y [F(2y-x) - F(x)]^{n-1} f(x)dx = n \int_0^{\delta_n} + n \int_{\delta_n}^{\varepsilon_n} + n \int_{\varepsilon_n}^y = I_1 + I_2 + I_3.$$

Estimate I_1:

$$I_1 = n \int_0^{\delta_n} [F(2y-x) - F(x)]^{n-1} f(x)dx \leqq$$

$$\leqq n \int_0^{\delta_n} [1 - F(x)]^{n-1} f(x)dx = \left[1 - \left(1 - F(\delta_n)\right)^n\right].$$

Estimate I_3:

$$I_3 = n \int_{\varepsilon_n}^y [F(2y-x) - F(x)]^{n-1} f(x)dx \leqq$$

$$\leqq n \int_{\varepsilon_n}^y [1 - F(x)]^{n-1} f(x)dx = O(1)\,[1 - F(\varepsilon_n)]^n.$$

Estimate I_2:

$$I_2=n\int_{\delta_n}^{\varepsilon_n}[F(2y-x)-F(x)]^{n-1}\,f(x)dx\geqq n\int_{\delta_n}^{\varepsilon_n}[F(2y-\varepsilon_n)-F(x)]^{n-1}\,f(x)dx=$$

$$=[F(2y-\varepsilon_n)-F(\delta_n)]^n-[F(2y-\varepsilon_n)-F(\varepsilon_n)]^n=$$

$$=F^n(2y-\varepsilon_n)\Big(1-\frac{F(\delta_n)}{F(2y-\varepsilon_n)}\Big)^n+O(1)\,[1-F(\varepsilon_n)]^n\,.$$

On the other hand

$$n\int_{\delta_n}^{\varepsilon_n}[F(2y-x)-F(x)]^{n-1}\,f(x)dx\leqq n\int_{\delta_n}^{\varepsilon_n}[F(2y-\delta_n)-F(x)]^{n-1}\,f(x)dx=$$

$$=F^n(2y-\delta_n)\Big(1-\frac{F(\delta_n)}{F(2y-\delta_n)}\Big)^n+O(1)\,[1-F(\varepsilon_n)]^n\,.$$

Summarizing the above estimates we obtain

$$b\geqq O(1)\,[1-F(\varepsilon_n)]^n+O(1)[1-(1-F(\delta_n))^n]+ \tag{6}$$
$$+F^n(2y-\varepsilon_n)\Big(1-\frac{F(\delta_n)}{F(2y-\varepsilon_n)}\Big)^n,$$

further

$$b\leqq O(1)\,[1-F(\varepsilon_n)]^n+O(1)[1-(1-F(\delta_n))^n]+ \tag{7}$$
$$+F^n(2y-\delta_n)\Big(1-\frac{F(\delta_n)}{F(2y-\delta_n)}\Big)^n.$$

Now choose ε_n and δ_n. Let ε_n and δ_n such that $(1-F(\varepsilon_n))^n\to 0$ as $n\to+\infty$ and $(1-F(\delta_n))^n\to 1$ as $n\to+\infty$. Because $\varepsilon_n\to 0$, $n\to+\infty$, hence $F(\varepsilon_n)$, $F(\delta_n)\to 0$ as $n\to+\infty$. New let $F(\varepsilon_n)=\alpha(n)/n$, where $0<\alpha(n)$ tends to infinity arbitrarily slowly and $\alpha(n)=o(n)$, further $F(\delta_n)=1/n\beta(n)$, where $0<\beta(n)$ tends to infinity arbitrarily slowly. Then

$$(1-F(\varepsilon_n))^n=\Big(1-\frac{\alpha(n)}{n}\Big)^n=O(1)e^{-\alpha(n)},$$

$$(1-F(\delta_n))^n=\Big(1-\frac{1}{n\beta(n)}\Big)^n=1+O(1)\frac{1}{\beta(n)},$$

further

$$\Big(1-\frac{F(\delta_n)}{F(2y-\varepsilon_n)}\Big)^n=1+O(1)\frac{1}{\beta(n)},$$

$$\left(1-\frac{F(\delta_n)}{F(2y-\delta_n)}\right)^n = 1+O(1)\frac{1}{\beta(n)},$$

namely, we have seen in (a) that for $n \geqq n_0(F,b)$ we have $y > 1/4$, hence $F(1/4) \leqq F(2y-\varepsilon_n) \leqq 1$.

Using the above estimates we obtain from (6) and (7)

$$b \geqq O(1)e^{-\alpha(n)} + O(1)\frac{1}{\beta(n)} + F^n(2y-\varepsilon_n)\left(1+O(1)\frac{1}{\beta(n)}\right),$$

further

$$b \leqq O(1)e^{-\alpha(n)} + O(1)\frac{1}{\beta(n)} + F^n(2y-\delta_n)\left(1+O(1)\frac{1}{\beta(n)}\right),$$

hence

$$\sqrt[n]{b}\left(1+O\left(\frac{e^{-\alpha(n)}}{n}\right)+O\left(\frac{1}{n\beta(n)}\right)\right) \geqq F(2y-\varepsilon_n),$$

$$\sqrt[n]{b}\left(1+O\left(\frac{e^{-\alpha(n)}}{n}\right)+O\left(\frac{1}{n\beta(n)}\right)\right) \geqq F(2y-\delta_n).$$

(ii) The case $F=[0,\infty]$. $y > 0$ because $f(x) = 0$ if $x < 0$. The same method as in the proof of Theorem A gives that $y \to +\infty$ $(n \to \infty)$. After this we can repeat the proof of (i) hence we omit the details.

The proof of the Theorem is complete.

Applications

I. Consider the case $f(x) = \frac{1}{\sqrt{2\pi}x}e^{-\log^2 x/2}$: "lognormal density function". It is well-known that

$$\int_x^\infty e^{-t^2/2}dt = e^{-x^2/2}(x^{-1}-x^{-3}+O(x^{-5})), \quad x > 1.$$

Using this fact and taking into account that $y \to +\infty$ $(n \to \infty)$ we get

$$F(2y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\log 2y} e^{-t^2/2}dt = 1-\frac{1}{\sqrt{2\pi}}\int_{\log 2y}^{\infty} e^{-t^2/2}dt =$$

$$= 1-\frac{e^{-\log^2 2y/2}}{\sqrt{2\pi}}\left(\frac{1}{\log 2y}+O(\log^{-3} y)\right).$$

On the other hand

$$\sqrt[n]{b} = 1 + \frac{\log b}{n} + O(n^{-2}).$$

We obtain from (2) without giving the details,

$$\log 2y \geqq \sqrt{2\log n}\left(1 + O\left(\frac{\log\log n}{\log n}\right)\right).$$

Using this estimate we obtain, applying Theorem A,

$$\frac{e^{-\log^2 2y/2}}{\sqrt{2\pi}\log 2y}(1 + O(\log^{-1} n)) = \frac{\log 1/b}{n}(1 + O(n^{-1})).$$

Hence

(8)

$$\log^2 y + \log 2\pi + 2\log\log 2y + O(\log^{-1} n) = -2\log\log 1/b + 2\log n + (n^{-1}).$$

We are looking for $\log 2y$ in the form $\log 2y = \sqrt{\log 2n}(1 + \varphi(n))$ where $\varphi(n) \to 0$ as $n \to \infty$. We obtain from (8):

$$\log\log n + \varphi(n)(4\log n + 2) + \varphi^2(n)2\log n =$$

$$= -2\log\log\frac{1}{b} - \log 2\pi - \log 2 + O(\log^{-1} n) + O(\varphi^2(n)).$$

Choose $\varphi(n)$ so that

$$\varphi^2(n)2\log n + \varphi(n)(4\log n + 2) + \log\log n + 2\log\log\frac{1}{b} + \log 4\pi = 0.$$

From this we obtain

$$\varphi(n) = -\frac{\log\log n}{4\log n} - \frac{2\log\log 1/b + \log 4\pi}{4\log n} + O\left(\left(\frac{\log\log n}{\log n}\right)^2\right).$$

Hence

$$2y = e^{\sqrt{2\log n}}e^{\sqrt{2\log n}\varphi(n)} = e^{\sqrt{2\log n}}\left(1 + \sqrt{2\log n}\varphi(n) + O(\log n\varphi^2(n))\right) =$$

$$= e^{\sqrt{2\log n}}\left(1 - \frac{\log\log n}{2\sqrt{2\log n}} - \frac{2\log\log 1/b + \log 4\pi}{2\sqrt{2\log n}} + O\left(\frac{(\log\log n)^2}{\log n}\right)\right),$$

i.e.

$$\frac{Q_{3/4} - Q_{1/4}}{2} = e^{\sqrt{2\log n}}\left(\frac{\log\log 4 - \log\log 3/4}{4\sqrt{2\log n}} + O\left(\frac{(\log\log n)^2}{\log n}\right)\right).$$

II. Now consider the case $f(x) = \lambda e^{-\lambda x}$, $x > 0$, $\lambda > 0$: "exponential density function".

In this case $F(x) = 1 - e^{-\lambda x}$ and from (2) we obtain

$$y \geqq \frac{\log n}{2\lambda}(1 + O(\log^{-1} n))$$

and so

$$\max_{2y-\varepsilon_n \leqq t \leqq 2y} f(t) = O(n^{-1}).$$

Applying Theorem B we get

$$F(2y) = \sqrt[n]{b} + O(\varepsilon_n/n) + O(e^{-\lambda_n \varepsilon_n})$$

and hence we obtain in the case of $\varepsilon_n = \frac{2\log n}{\lambda_n}$:

$$F(2y) = 1 + \frac{\log b}{n} + O\Big(\frac{\log n}{n^2}\Big),$$

hence

$$y = \frac{\log n}{2\lambda} - \frac{\log\log 1/b}{2\lambda} + O\Big(\frac{\log n}{n}\Big),$$

i.e.

$$\frac{Q_{3/4} - Q_{1/4}}{2} = \frac{\log\log 4 - \log\log 4/3}{4\lambda} + O\Big(\frac{\log n}{n}\Big).$$

III. Now consider the case

$$f(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1}, \qquad 0 < x < 1; p, q > 0:$$

"beta density function". We need the following lemma.

LEMMA. *Let* (x_n) *be such that* $\lim x_n = 1, 0 < x_n < 1$. *Then*

$$\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^{x_n} t^{p-1}(1-t)^{q-1} dt = 1 - \frac{\Gamma(p+q)}{q\Gamma(p)\Gamma(q)}(1-x_n)^q + O(1)(1-x_n)^{q+1},$$

where $O(1)$ *is an effective constant depending only on* p *and* q.

PROOF. Obviously,

$$\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^{x_n} t^{p-1}(1-t)^{q-1} dt = 1 - \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_{x_n}^1 t^{p-1}(1-t)^{q-1} dt,$$

further

$$\int_{x_n}^1 t^{p-1}(1-t)^{q-1} dt = \int_{x_n}^1 (t^{p-1} - 1)(1-t)^{q-1} dt + \int_{x_n}^1 (1-t)^{q-1} dt = I_1 + I_2.$$

We have

$$I_2 = \int_{x_n}^{1} (1-t)^{q-1}dt = \frac{1}{q}\Big(1-x_n\Big)^q,$$

$$I_1 = -\int_{x_n}^{1} (1-t^{p-1})(1-t)^{q-1}dt = -\int_{x_n}^{1} \frac{1-t^{p-1}}{1-t}(1-t)^q dt.$$

Obviously,

$$\lim_{t\to 1-0} \frac{1-t^{p-1}}{1-t} = p-1 \quad \text{if} \quad p \neq 1$$

and

$$\frac{1-t^{p-1}}{1-t} = 0 \quad \text{if} \quad p = 1$$

hence

$$\frac{1-t^{p-1}}{1-t} = O(1),$$

where $O(1)$ is an effective constant depending only on p. Consequently,

$$I_1 = O(1)\int_{x_n}^{1} (1-t)^q dt = O(1)(1-x_n)^{q+1}.$$

The Lemma is proved.

Because $F(\delta_n) < F(\varepsilon_n) = \frac{\alpha(n)}{n} \to 0$ and F is continuous in 0 hence $0 < \delta_n < \varepsilon_n \to 0$ $(n \to \infty)$. According to $F(z) \asymp z^p$, $0 < z < 1/2$ we have $F(\varepsilon_n) \asymp \varepsilon_n^p$, $F(\delta_n) \asymp \delta_n^p$.

Let

$$\varepsilon_n := \Big(\frac{\alpha(n)}{n}\Big)^{1/p}, \quad \delta_n := \Big(\frac{1}{n\beta(n)}\Big)^{1/p}.$$

Because $\sqrt[n]{b} < 1$ and $\lim_{n\to\infty} \sqrt[n]{b} = 1$ $(0 < b < 1)$, hence it follows from (4) and (5) that $2y - \varepsilon_n < 1$ further $y \to 1/2$ $(n \to \infty)$.

Applying the Lemma and taking into account (4) and (5) we get

$$1 + \frac{\log b}{n} + O(n^{-2}0) + O\Big(\frac{e^{-\alpha(n)}}{n}\Big) + O\Big(\frac{1}{n\beta(n)}\Big) \geqq$$

$$\geqq 1 - \frac{\Gamma(p+q)}{q\Gamma(p)\Gamma(q)}(1-2y+\varepsilon_n)^q + O((1-2y+\varepsilon_n)^{q+1}),$$

$$1 + \frac{\log b}{n} + O(n^{-2}) + O\Big(\frac{e^{-\alpha(n)}}{n}\Big) + O\Big(\frac{1}{n\beta(n)}\Big) \leqq$$

$$\leqq 1 - \frac{\Gamma(p+q)}{q\Gamma(p)\Gamma(q)}(1-2y+\delta_n)^q + O((1-2y+\delta_n)^{q+1}),$$

i.e.

$$\frac{\log\frac{1}{b}}{n}\frac{q\Gamma(p)\Gamma(q)}{\Gamma(p+q)}\left[1+O(n^{-1})+O(e^{-\alpha(n)})+O\left(\frac{1}{\beta(n)}\right)\right] \leqq \tag{9}$$
$$\leqq (1-2y+\varepsilon_n)^q[1+O(1-2y+\varepsilon_n)],$$

$$\frac{\log\frac{1}{b}}{n}q\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}\left[1+O(n^{-1})+O(e^{-\alpha(n)})+O\left(\frac{1}{\beta(n)}\right)\right] \geqq \tag{10}$$
$$\geqq (1-2y+\delta_n)^q[1+O(1-2y+\delta_n)].$$

According to $O(1-2y+\delta_n)=o(1)$ we obtain from (10) $1-2y+\delta_n = O(1)(\frac{1}{n})^{1/q}$. Using this and taking into account (10) again we get

$$\frac{\log\frac{1}{b}}{n}\frac{q\Gamma(p)\Gamma(q)}{\Gamma(p+q)}\left[1+O(n^{-1})+O\left(1/\sqrt[q]{n}\right)+O(e^{-\alpha(n)})+O\left(\frac{1}{\beta(n)}\right)\right] \geqq$$

$$\geqq (1-2y+\delta_n)^q.$$

Choose $\beta(n)=e^n$. Then it follows

$$y \geqq \frac{1}{2} - \left[\left(\log\frac{1}{b}\right)\frac{q\Gamma(p)\Gamma(q)}{\Gamma(p+q)}\right]^{1/q}\frac{1}{\sqrt[q]{n}}\times \tag{11}$$
$$\times\left[1+O(n^{-1})+O\left(1/\sqrt[q]{n}\right)+O(e^{-\alpha(n)})\right].$$

Now we give an upper estimate for y. Because

$$1-2y+\varepsilon_n = 1-2y+\delta_n+(\varepsilon_n-\delta_n) = O\left(1/\sqrt[q]{n}\right)+O\left(\left(\frac{\alpha(n)}{n}\right)^{1/p}\right)$$

hence from (9) we get

$$\frac{\log\frac{1}{b}}{n}\frac{q\Gamma(p)\Gamma(q)}{\Gamma(p+q)}\left[1+O(n^{-1})+O\left(1/\sqrt[q]{n}\right)+O\left(\left(\frac{\alpha(n)}{n}\right)^{1/p}\right)+\right.$$

$$\left.+O(e^{-\alpha(n)})+O\left(\frac{1}{\beta(n)}\right)\right] \leqq (1-2y+\varepsilon_n)^q,$$

i.e.

$$y \leqq \frac{1}{2} - \left[\left(\log\frac{1}{b}\right)\frac{q\Gamma(p)\Gamma(q)}{\Gamma(p+q)}\right]^{1/q}\frac{1}{\sqrt[q]{n}}\times \tag{12}$$

$$\times\left[1+O(n^{-1})+O\left(1/\sqrt[q]{n}\right)+O\left(\left(\frac{\alpha(n)}{n}\right)^{1/p}\right)+O(e^{-\alpha(n)})\right]+O\left(\left(\frac{\alpha(n)}{n}\right)^{1/p}\right).$$

From (11) and (12) we obtain

$$y=\frac{1}{2}-\left[\left(\log\frac{1}{b}\right)\frac{q\Gamma(p)\Gamma(q)}{\Gamma(p+q)}\right]^{1/q}\frac{1}{\sqrt[q]{n}}\times$$

$$\times\left[1+O(n^{-1})+\left(1/\sqrt[q]{n}\right)+O\left(\left(\frac{\alpha(n)}{n}\right)^{1/p}\right)+O(e^{-\alpha(n)})\right]+O\left(\left(\frac{\alpha(n)}{n}\right)^{1/p}\right),$$

consequently,

$$\frac{Q_{3/4}-Q_{1/4}}{2}=\frac{(\log 4)^{1/q}-(\log\frac{4}{3})^{1/q}}{2}\left(\frac{q\Gamma(p)\Gamma(q)}{\Gamma(p+q)}\right)^{1/q}\frac{1}{\sqrt[q]{n}}\times$$

$$\times\left[1+O(n^{-1})+O\left(1/\sqrt[q]{n}\right)+O\left(\left(\frac{\alpha(n)}{n}\right)^{1/p}\right)+O(e^{-\alpha(n)})\right]+O\left(\left(\frac{\alpha(n)}{n}\right)^{1/p}\right).$$

If $q>p$, then let $\alpha(n)=\log n$. Then we obtain

$$\frac{Q_{3/4}-Q_{1/4}}{2}=\frac{(\log 4)^{1/q}-(\log\frac{4}{3})^{1/q}}{2}\left(\frac{q\Gamma(p)\Gamma(q)}{\Gamma(p+q)}\right)^{1/q}\frac{1}{\sqrt[q]{n}}\times$$

$$\times\left[1+O(n^{-1})+O\left(1/\sqrt[q]{n}\right)+O((\log n)^{1/p}n^{\frac{1}{q}-\frac{1}{p}})\right].$$

It follows that in the case $q<2$ our estimate is better, in the case $q=2$ our estimate is equally good, further in the case $q>2$ it is worse than the arithmetic mean.

If $q\leqq p$, then

$$\frac{Q_{3/4}-Q_{1/4}}{2}=O\left(\left(\frac{\alpha(n)}{n}\right)^{1/p}\right),$$

which is better at $p<2$ than the estimate given by the arithmetic mean.

PROBLEM. What is true if $q\leqq p$?

REMARK 1. In [1] (VII.) we have considered the probability density functions

$$f(x)=d_2\frac{x^{\varepsilon}}{(\alpha x^{\beta}+\gamma)^{\frac{\delta+\varepsilon+2}{\beta}}}\qquad(x\geqq x_1>0),\quad d_2>0,$$

where

$$\alpha=0,\quad\beta>0,\quad\gamma>0,\quad\delta\in R,\quad\varepsilon<-1,$$

or

$$\alpha>0,\quad\beta>0,\quad\gamma\in R,\quad\delta>-1,\quad\varepsilon\in R,$$

or

$$\alpha \in R, \quad \beta < 0, \quad \gamma > 0, \quad \delta \in R, \quad \varepsilon < -1.$$

The reason of this is the following. Let ξ and η be independent random variables with density functions

$$h(x) = \frac{\beta \alpha^{\frac{\varepsilon+1}{\beta}}}{\Gamma(\frac{\varepsilon+1}{\beta})} x^{\varepsilon} e^{-\alpha x^{\beta}}, \qquad x > 0; \quad \alpha, \beta > 0; \quad \varepsilon > -1$$

and

$$g(x) = \frac{\beta \gamma^{\frac{\delta+1}{\beta}}}{\Gamma(\frac{\delta+1}{\beta})} x^{\delta} e^{-\gamma x^{\beta}}, \qquad x > 0; \quad \gamma, \beta > 0; \quad \delta > -1,$$

respectively. Then $f(x)$ is the density function of the random variable ξ/η. Namely, according to [3] (p. 105) the density function $s(x)$ of ξ/η is

$$s(x) = \frac{\beta \alpha^{\frac{\varepsilon+1}{\beta}} \beta \gamma^{\frac{\delta+1}{\beta}}}{\Gamma(\frac{\varepsilon+1}{\beta}) \Gamma(\frac{\delta+1}{\beta})} \int_0^{\infty} y x^{\varepsilon} y^{\varepsilon} e^{-\alpha x^{\beta} y^{\beta}} y^{\delta} e^{-\gamma y^{\beta}} dy.$$

Calculating the integral on the right-hand side we obtain

$$s(x) = \frac{\beta \alpha^{\frac{\varepsilon+1}{\beta}} \gamma^{\frac{\delta+1}{\beta}} \Gamma(\frac{\delta+\varepsilon+2}{\beta})}{\Gamma(\frac{\varepsilon+1}{\beta}) \Gamma(\frac{\delta+1}{\beta})} x^{\varepsilon} (\alpha x^{\beta} + \gamma)^{-\frac{\delta+\varepsilon+2}{\beta}},$$

i.e. $s(x)$ and $f(x)$ are the same type. Taking into account $\int_{-\infty}^{\infty} s(x) dx = 1$, ($s(x)$ is a density function) we get

$$\int_0^{\infty} x^{\varepsilon} (\alpha x^{\beta} + \gamma)^{-\frac{\delta+\varepsilon+2}{\beta}} dx = \frac{1}{\beta} \alpha^{-\frac{\varepsilon+1}{\beta}} \gamma^{-\frac{\delta+1}{\beta}} \frac{\Gamma(\frac{\varepsilon+1}{\beta}) \Gamma(\frac{\delta+1}{\beta})}{\Gamma(\frac{\delta+\varepsilon+2}{\beta})}. \tag{13}$$

A special case of this formula is given in [4] (p. 292, 3. 241 (4)). As illustrations we give two special cases of this formula.

(a) Let $\varepsilon = 0; \quad \alpha, \gamma = 1; \quad \frac{\delta+2}{\beta} = 1, \quad \beta > 1$. Then we get from (13)

$$\int_0^{\infty} \frac{1}{x^{\beta} + 1} dx = \frac{\Gamma(\frac{1}{\beta}) \Gamma(1 - \frac{1}{\beta})}{\beta}.$$

The left-hand side is elementarily integrable (see [5], II, 459, 484.) and we obtain

$$\frac{\pi}{\sin \frac{\pi}{\beta}} = \Gamma\left(\frac{1}{\beta}\right) \Gamma\left(1 - \frac{1}{\beta}\right),$$

i.e.

$$\frac{\pi}{\sin \pi x} = \Gamma(x)\Gamma(1-x), \quad 0 < x < 1. \tag{14}$$

Taking into account the periodicity of the sine function and the equation $\Gamma(x+1) = x\Gamma(x)$ we obtain (14) for any $x \in \mathbf{R} \setminus \mathbf{Z}$.

(b) Let $\alpha, \beta, \gamma = 1$; $\varepsilon, \delta > -1$. Then we get from (13)

$$\int_0^\infty \frac{x^\varepsilon}{(1+x)^{\delta+\varepsilon+2}} dx = \frac{\Gamma(\varepsilon+1)\Gamma(\delta+1)}{\Gamma(\delta+\varepsilon+2)},$$

and hence, by the substitution $\frac{x}{1+x} = y$,

$$\int_0^1 y^\varepsilon (1-y)^\delta dy = \frac{\Gamma(\varepsilon+1)\Gamma(\delta+1)}{\Gamma(\delta+\varepsilon+2)} \qquad (\varepsilon, \delta > -1).$$

The integral on the left-hand side is widely known, this is Euler's beta function.

REMARK 2. Next we show that our formula (13) gives an elemantary proof for a Titchmarsh formula for the Riemann function. To this let

$$\alpha > 0, \ \beta \geqq 0 \ \text{ such that } \ -1 < \frac{1}{\alpha} - \frac{1}{\beta} < 1 \ \text{ and}$$

$$f(t) = \int_0^\infty z^{\frac{1}{\alpha}-\frac{\beta}{\alpha}-1} \sin tz \, dz, \quad t \in \mathbf{R}. \tag{15}$$

According to our assumptions, the integral exists. The substitution $z = u/t$ shows that $f(t) = t^{\frac{\beta}{\alpha}-\frac{1}{\alpha}} f(1)$. Multiplying both sides of (15) by $e^{-pt}(p > 0)$, and integrating with respect to t from 0 to $+\infty$ we obtain

$$f(1) \int_0^\infty t^{\frac{\beta}{\alpha}-\frac{1}{\alpha}} e^{-pt} dt = \int_0^\infty \Big(\int_0^\infty z^{\frac{1}{\alpha}-\frac{\beta}{\alpha}-1} \sin tz \, dz \Big) e^{-pt} dt,$$

and by Fubini's theorem we get

$$f(1) \Big(\frac{1}{p}\Big)^{\frac{\beta}{\alpha}-\frac{1}{\alpha}+1} \Gamma\Big(\frac{\beta}{\alpha} - \frac{1}{\alpha} + 1\Big) = \int_0^\infty \frac{1}{z^{\frac{\beta}{\alpha}-\frac{1}{\alpha}+1}} \frac{z}{p^2+z^2} dz. \tag{16}$$

(Using (13) in the case $p=1$, $\alpha=\gamma=1$, $\beta=2$, $\varepsilon=\frac{1}{\alpha}-\frac{\beta}{\alpha}$, $\delta=\frac{\beta}{\alpha}-\frac{1}{\alpha}$ we obtain

$$f(1)=\int_0^\infty z^{\frac{1}{\alpha}-\frac{\beta}{\alpha}-1}\sin z\,dz=\frac{\pi}{2}\frac{1}{\Gamma(\frac{\beta}{\alpha}-\frac{1}{\alpha}+1)}\frac{1}{\sin\frac{\pi}{2}(\frac{1}{\alpha}-\frac{\beta}{\alpha}+1)}. \tag{17}$$

Let $s=\frac{\beta}{\alpha}-\frac{1}{\alpha}+1$ and $\beta>1$, i.e. $1<s<2$. Then using (16) and (17) we get

$$\left(\frac{1}{p}\right)^s=\frac{2}{\pi}\sin\frac{\pi s}{2}\int_0^\infty\frac{1}{z^{s-1}}\frac{1}{p^2+z^2}dz,$$

and hence, summing in p from 1 to ∞ we obtain

$$\zeta(s)=\frac{2}{\pi}\sin\frac{\pi s}{2}\int_0^\infty\frac{1}{z^{s-1}}\sum_{p=1}^\infty\frac{1}{p^2+z^2}dz.$$

It is well-known that

$$\pi\cot\pi z=\frac{1}{z}+\frac{2z}{z^2-1^2}+\cdots+\frac{2z}{z^2-k^2}+\cdots,\quad z\in\mathbf{C}\setminus\mathbf{Z}$$

hence

$$\sum_{p=1}^\infty\frac{1}{p^2+z^2}=-\frac{\pi\cot\pi iz+\frac{i}{z}}{2iz}=\frac{\pi}{2z}\frac{e^{\pi z}+e^{-\pi z}}{e^{\pi z}-e^{-\pi z}}-\frac{1}{2z^2}=\frac{1}{2z}\left(\pi\coth\pi z-\frac{1}{z}\right),$$

consequently

$$\zeta(s)=\frac{1}{\pi}\sin\frac{\pi s}{2}\int_0^\infty\frac{1}{z^s}\left(\pi\coth\pi z-\frac{1}{z}\right)dz,\quad 1<s<2.$$

Consider the decomposition

$$\int_0^\infty\frac{1}{z^s}\left(\pi\coth\pi z-\frac{1}{z}\right)dz=\int_0^1\frac{1}{z^s}\left(\pi\coth\pi z-\frac{1}{z}\right)dz+$$

$$+\int_1^\infty\frac{1}{z^s}\left(\pi\coth\pi z-\frac{1}{z}\right)dz=I_1+I_2.$$

Using the well-known expansion

$$\coth x=\frac{1}{x}+\sum_{n=1}^\infty\frac{2^{2n}B_{2n}}{(2n)!}x^{2n-1},\quad 0<x<\pi,$$

(here B_{2n} denotes the $2n-$th Bernoulli number)

$$I_1 = \sum_{n=1}^{\infty} \frac{2^{2n}\pi^{2n}B_{2n}}{(2n)!}\frac{1}{2n-s},$$

and applying this we get

$$\zeta(2m) = (-1)^{m-1}\frac{(2\pi)^{2m}}{2(2m)!}B_{2m},$$

i.e.

$$I_1 = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{2n-s}\zeta(2n).$$

On the other hand

$$I_2 = \int_1^{\infty}\Big(\frac{\pi}{z^s} + \frac{2\pi}{e^{2\pi z}-1}\frac{1}{z^s} - \frac{1}{z^{s+1}}\Big)dz = \frac{\pi}{s-1} - \frac{1}{s} + 2\pi\int_1^{\infty}\frac{1}{z^s(e^{2\pi z}-1)}dz.$$

Summarizing our results we obtain

$$(18)\quad \zeta(s) = \frac{1}{\pi}\sin\frac{\pi s}{2}\Big(\frac{\pi}{s-1} - \frac{1}{s} + \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{2n-s}\zeta(2n) + 2\pi\int_1^{\infty}\frac{1}{z^s(e^{2\pi z}-1)}dz\Big),$$

where $1 < s < 2$.

Here

$$\int_1^{\infty}\frac{1}{z^s(e^{2\pi z}-1)}dz < c(s) < \infty, \quad s \in \mathbf{R},$$

which means that (18) holds for every $s \in \mathbf{R}\setminus\{1\}$. The function on the right-hand side of (18) is analytical in $\mathbf{R}\setminus\{1,0,2,4,\dots,2m,\dots\}$ it has a pole in $s=1$, analytically continuable to $\mathbf{R}\setminus\{1\}$, in real sense. According to the unicity of the analytical continuation we obtain that (18) holds for every $s \in \mathbf{C}\setminus\{1\}$.

REFERENCES

[1] Joó, I. and Szabó, S., On the estimate $(x_{\min}+x_{\max})/2$, *Studia Sci. Math. Hungar.* **27** (1992), 409–432.

[2] Csernyák, L., Flanken-Diagnostik, *Publ. Technical Univ. for Heavy Industry*, Series A, Miskolc.

[3] Prékopa, A., *Valószínűségelmélet* [Probability theory], Tankönyvkiadó, Budapest, 1962 (in Hungarian).

[4] GRADSHTEĬN, I. S. and RYZHIK, I. M., *Table of integrals, series, and products,* Corrected and enlarged edition, Academic Press, New York–Toronto–London, Ont., 1980. *MR* **81g:** 33001

[5] FIKHTENGOL'TS, G. M., *Kurs differentsial'nogo i integral'nogo ischisleniya* [Course in differential and integral calculus], vols. I, II, and III, Gostekhizdat, Moscow and Leningrad, 1947–1949.

(Received February 27, 1990)

MTA MATEMATIKAI KUTATÓINTÉZETE
P. O. BOX 127
H–1364 BUDAPEST
HUNGARY

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
ALKALMAZOTT ANALÍZIS TANSZÉK
MÚZEUM KRT. 6–8
H–1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 339–348

MAL'TSEV FUNCTIONS ON SMALL ALGEBRAS

I. CHAJDA and G. CZÉDLI

Abstract

The following problem is considered. Given an n-element set A and a set L of permuting equivalences on A, does there exist a Mal'tsev function $A^3 \to A$ which is compatible with all members of L? The answer is negative in general when $n \geqq 25$, it remains open for $9 \leqq n \leqq 24$, and it is shown to be affirmative for $n \leqq 8$. Moreover, there is even a commutative Mal'tsev function when $n \leqq 8$.

Introduction and result

Given a set A, a function $p\colon A^3 \to A$ is called a Mal'tsev function if $p(x,y,y) = p(y,y,x) = x$ holds for any $x, y \in A$. If an algebra A has a Mal'tsev function $p\colon A^3 \to A$ which is compatible with all congruences of A then A is congruence permutable. However, the converse is not true in general (cf. Gumm [3]). The purpose of the present paper is to furnish the converse statement under the additional condition $|A| \leqq 8$. In order to obtain a somewhat stronger statement we formulate our result not only for algebras. Then it may be of some interest in studying intersections of certain maximal clones on a finite set with less than nine elements. A Mal'tsev function $p\colon A^3 \to A$ is called commutative if $p(x_{1\pi}, x_{2\pi}, x_{3\pi}) = p(x_1, x_2, x_3)$ holds for any $(x_1, x_2, x_3) \in A^3$ and any permutation $\pi\colon \{1,2,3\} \to \{1,2,3\}$.

THEOREM. *Let A be a set with $|A| \leqq 8$ and let L be a sublattice of the lattice of equivalences on A. Then the following three conditions are equivalent:*

(a) *the equivalences belonging to L permute, i.e., for any $\rho, \nu \in L$, $\rho \circ \nu = \nu \circ \rho$;*

(b) *there exists a Mal'tsev function $A^3 \to A$ which is compatible with any member of L;*

1980 *Mathematics Subject Classification* (1985 Revision). Primary 08B05.
Key words and phrases. Congruence permutability, Mal'tsev term, finite algebra.

The second author's work was partially supported by the Hungarian Foundation for Scientific Research Grant No. 1813.

Akadémiai Kiadó, Budapest

(c) *there is a commutative Mal'tsev function* $A^3 \to A$ *which is compatible with (any member of)* L.

Our method yielding the equivalence of (a) and (b) for $|A| \leqq 8$ is possibly applicable for $|A| = 9$ or $|A| = 10$ or even more. However, the length of the proof would grow rather fast with $|A|$ and we do not want to make it astronomically long. Another excuse for stopping at eight is that for $|A| = 9$ (a) and (c) are not equivalent. Really, if A is the square of the three element group and L is its congruence lattice then (a) holds but (c) does not (cf. Gumm [4, Thm. 3.2]).

While the equivalence of (a) and (b) is an open problem for $|A| \in \{9, 10, \ldots, 24\}$, they are not equivalent for $|A| \geqq 25$. Moreover, we have the following

OBSERVATION. *For any natural number* $n \geqq 25$ *there is an* n*-element algebra* A *such that* A *has permutable congruences but no Mal'tsev function* $A^3 \to A$ *is compatible with all congruences of* A.

PROOF. Starting from a five-element non-associative loop (cf. Gumm [4, Fig. 2.4]) Gumm constructed a twentyfive-element A with the required property in [3]. Suppose we already have an n-element algebra $A = (A, F)$ as required, then we construct an $(n+1)$-element algebra B in the following way. Put $B = A \cup \{w\}$ where $w \notin A$. For $f\colon A^k \to A$ in F define $f_B\colon B^k \to B$,

$$f_B(b_1, \ldots, b_k) = \begin{cases} f(b_1, \ldots, b_k) & \text{if } b_1, \ldots, b_k \in A \\ w & \text{otherwise.} \end{cases}$$

Further, for any $c \in A$, define $g_c\colon B \to B$ by

$$g_c(x) = \begin{cases} x & \text{if } x \neq w \\ c & \text{if } x = w. \end{cases}$$

Now put $B = (B, \{f_B : f \in F\} \cup \{g_c : c \in A\})$. Then for any nontrivial congruence α of B the block $[w]\alpha$ is a singleton and $\alpha|_A$ is a congruence of A. Thus the congruences of B permute. We can observe that any congruence of A is the restriction of a (unique) congruence of B. In particular, B has a congruence κ with exactly two blocks: A and $\{w\}$. Suppose B permits a compatible Mal'tsev function $p\colon B^3 \to B$. Then, for $x, y, z \in A$, $p(x, y, z)\,\kappa\, p(x, x, x) = x$ whence $p(x, y, z) \in A$. Therefore the restriction of p to A is a compatible Mal'tsev function on A, contradicting the induction hypothesis. Q.e.d.

PROOF OF THE THEOREM. The implication (b) $\Rightarrow$ (a) follows from the classical argument of Mal'tsev [5]. Namely, if $u, v \in A$, $\alpha, \beta \in L$ and $(u, v) \in \alpha \circ \beta$ then there is an element $w \in A$ with $u \alpha w \beta v$. If p is a compatible Mal'tsev function then

$$u = p(u, v, v)\,\beta\, p(u, w, v)\,\alpha\, p(u, u, v) = v$$

whence $(u, v) \in \beta \circ \alpha$. The implication (c) $\Rightarrow$ (b) being trivial we have to show only that (a) implies (c). This will need several preliminaries.

We will often consider diamonds (five-element non-distributive modular sublattices) in L; their elements will be denoted by $\omega, \alpha, \beta, \gamma, \iota$ such that $\omega -< \alpha -< \iota$, $\omega -< \beta -< \iota$, $\omega -< \gamma -< \iota$. The bottom and the top of L is denoted by 0 and 1, respectively.

Let $n \leqq 8$ and assume that (a) $\Rightarrow$ (c) for sets consisting of less than n elements. We fix an n-element set A and a permutable sublattice L of the equivalence lattice of A. We have to show the existence of a commutative Mal'tsev function which is compatible with L. A particular case is settled by the following

LEMMA 1. *If there exists a $\mu \in L\backslash\{0\}$ such that $\mu \leqq \omega$ holds for every diamond $\{\omega, \alpha, \beta, \gamma, \iota\}$ in L then we are done. (I.e., then there is a commutative Mal'tsev function which is compatible with L.)*

PROOF. The proof of this lemma borrows a lot of ideas from Pixley [6, p. 183]. By the induction hypothesis, there is a commutative Mal'tsev function $P_\mu : (A/\mu)^3 \to A/\mu$ preserving all ν/μ where $\mu \leqq \nu \in L$. For each $\lambda \in L$ we intend to define a commutative Mal'tsev function $p_\lambda : (A/\lambda)^3 \to A/\lambda$ preserving all ν/λ ($\lambda \leqq \nu \in L$) such that for any $\lambda_1 \leqq \lambda_2 \in L$

$$p_{\lambda_1}([x]\lambda_1, [y]\lambda_1, [z]\lambda_1) \subseteq p_{\lambda_2}([x]\lambda_2, [y]\lambda_2, [z]\lambda_2) \tag{1}$$

for any $x, y, z \in A$. Then we will be ready as $p_0 : A^3 \to A$ is what we are looking for.

Let us fix a linear order on A. First we define p_λ for $\lambda \geqq \mu$ as follows:

$$p_\lambda([x]\lambda, [y]\lambda, [z]\lambda) = \{t \in A : ([t]\mu, p_\mu([x]\mu, [y]\mu, [z]\mu)) \in \lambda/\mu\}.$$

Roughly speaking, this is $[p_\mu([x]\mu, [y]\mu, [z]\mu)]\lambda/\mu$ apart from the canonical correspondence between A/λ and $(A/\mu)/(\lambda/\mu)$. Then for $\lambda = \mu$ p_λ is just the previously defined p_μ. A routine calculation shows that p_λ is a commutative Mal'tsev function preserving all ν/λ ($\nu \geqq \lambda$) and (1) holds for $\mu \leqq \lambda_1 \leqq \lambda_2$.

Now we define p_λ for $\lambda \not\geqq \mu$ via a downward induction on the height of λ. (Note that L is a modular lattice, for its members permute.) Assume that $\lambda \not\geqq \mu$ and $p_{\lambda'}$ is already defined for each $\lambda' > \lambda$ such that the required properties, including (1), are satisfied for these λ'. Let $\nu_1, \ldots, \nu_k$ be the upper covers of λ and define p_λ as follows.

Let $p_\lambda([x]\lambda, [y]\lambda, [z]\lambda) = [a]\lambda$ where if two of the blocks $[x]\lambda, [y]\lambda$ and $[z]\lambda$ coincide then a is the first element in the remaining block. Otherwise let a be the first element in the intersection

$$\bigcap_{i=1}^{k} p_{\nu_i}([x]\nu_i, [y]\nu_i, [z]\nu_i). \tag{2}$$

(This will be shown nonempty later.)

Now if, e.g., $[x]\lambda = [y]\lambda$ then $[x]\nu_i = [y]\nu_i$ yields that $[z]\lambda$ is a subset of (2). Therefore a always belongs to the intersection (2). Thus p_λ is a commutative

Mal'tsev function. The property (1) extends to λ easily. Indeed, if $\lambda < \lambda_2$ then $\lambda -< \nu_i \leqq \lambda_2$ for some i and $p_\lambda([x]\lambda,[y]\lambda,[z]\lambda) = [a]\lambda \subseteq [a]\nu_i = p_{\nu_i}([x]\nu_i,[y]\nu_i,[z]\nu_i) \subseteq p_{\lambda_2}([x]\lambda_2,[y]\lambda_2,[z]\lambda_2)$. Using a routine calculation or referring to Pixley's proof [6, p. 183] we can see that p_λ is compatible with all ν/λ, $\nu \geqq \lambda$.

Now we set off to prove that (2) is not empty. We claim that

$$\prod_{i=1}^{j-1}(\nu_j + \nu_i) = \nu_j \quad \text{for } 2 < j < k. \tag{3}$$

(Here and in the sequel $+$ and $\cdot$ stand for the lattice operations join and meet, respectively.) Since the role of the ν_l $(1 \leqq l \leqq k)$ is symmetric, it suffices to deal with $j = 3$. Then (3) turns into $(\nu_3 + \nu_1)(\nu_3 + \nu_2) = \nu_3$. It belongs to the folklore of lattice theory that if $(x_3 + x_1)(x_3 + x_2) > x_3$ for distinct atoms x_1, x_2, x_3 in a modular lattice M then $\{x_1, x_2, x_3\}$ generates a diamond with bottom 0_M and top $x_3 + x_1$. Indeed, by the properties of the height function (cf., e.g., Grätzer [2]), $x_3 + x_1$ and $x_3 + x_2$ are of height two and so is their meet by the assumption. Thus $x_3 + x_1 = x_3 + x_2$. Since $x_1 + x_2$ is of height two either and $x_1 + x_2 \leqq (x_3 + x_1) + (x_3 + x_2) = x_3 + x_1$, $x_1 + x_2 = x_3 + x_1$. Since L is modular (cf., e.g., Grätzer [2, Thm. IV.4.10 and the remark after its proof]), we can apply the above observation for the interval $[\lambda, 1]$. Therefore $(\nu_3 + \nu_1)(\nu_3 + \nu_2) = \nu_3$ as otherwise λ would be the bottom of a diamond in spite of $\lambda \not\geqq \mu$.

The next step is to show

(4) If $a_i \in A$ and for all $i, j \leqq k$ $(a_i, a_j) \in \nu_i + \nu_j$ then there exists an element $b \in A$ such that $(a_i, b) \in \nu_i$ for all $i \leqq k$.

Indeed, this says nothing for $k = 1$ and follows from $\nu_1 + \nu_2 = \nu_1 \circ \nu_2$ for $k = 2$. If we have found an element b already such that $(a_i, b) \in \nu_i$ for $i = 1, 2, \ldots, j$ $(2 \leqq j < k)$ then $(b, a_{j+1}) \in \nu_i \circ (\nu_i + \nu_{j+1}) = \nu_i + \nu_{j+1}$ for all $i \leqq j$ and (3) yields $(b, a_{j+1}) \in \prod_{i \leqq j}(\nu_{j+1} + \nu_i) = \nu_{j+1}$. Therefore $(a_i, b) \in \nu_i$ holds for all $i \leqq k$.

Now, returning to (2), pick an element a_i in $p_{\nu_i}([x]\nu_i,[y]\nu_i,[z]\nu_i)$, $i = 1, 2, \ldots, k$. By the induction hypothesis made on λ, for $i, j \leqq k$ we have

$$a_i \in p_{\nu_i}([x]\nu_i,[y]\nu_i,[z]\nu_i) \subseteq$$
$$\subseteq p_{\nu_i+\nu_j}([x](\nu_i+\nu_j),[y](\nu_i+\nu_j),[z](\nu_i+\nu_j)),$$

and a_j belongs there, too. Hence $(a_i, a_j) \in \nu_i + \nu_j$. Now (4) supplies us with an element b such that $b\nu_i a_i$ for all i. I.e., $b \in [a_i]\nu_i = p_{\nu_i}([x]\nu_i,[y]\nu_i,[z]\nu_i)$. This b belongs to the intersection (2). Q.e.d.

Let us call an element $\mu \in L$ semicentral if $\mu \circ \nu = \mu \cup \nu$ (set theoretic union) holds for every $\nu \in L$. (Note that $\mu \circ \nu = \mu + \nu$ by permutability.)

LEMMA 2. *If there exists a semicentral $\mu \in L\setminus\{0,1\}$ then we are done.*

PROOF. Let $B_1, B_2, \ldots, B_t$ be the μ-blocks. Since μ is not in $\{0,1\}$, we have $t < n$ and $|B_i| < n$ for all i. Observe that the restrictions of members of L to B_i permute. Indeed, if $\rho, \nu \in L$, $a, b, c \in B_i$, $a\rho c$ and $c\nu b$ then there is a $d \in A$ with $a\nu d\rho b$. If $d \notin B_i$ then $(c,d) \in \mu \circ \nu = \mu \cup \nu$ yields $c\nu d$, whence $a\nu b$ by transitivity. Therefore $a\nu b\rho b$, showing that the restrictions of ν and ρ to B_i permute. By the induction hypothesis on $|A|$ there is a commutative Mal'tsev function $p_i\colon B_i^3 \to B_i$ preserving the restrictions of members of L for each i, $1 \leqq i \leqq t$. Similarly, there is a Mal'tsev function $p_\mu\colon (A/\mu)^3 \to A/\mu$ preserving all the ρ/μ, $\mu \leqq \rho \in L$. Now let us fix an element $b_i \in B_i$ for each i, $1 \leqq i \leqq t$. For $x, y, z \in A$ let $B_k = B_k(x,y,z)$ be $p_\mu([x]\mu, [y]\mu, [z]\mu)$ and define $u = p(x,y,z)$ as follows:

(α) if x, y, z belong to the same μ-block B_j then $u = p_j(x,y,z)$ (note that $j = k$);

(β) if $|\{x,y,z\} \cap B_k| = 1$ then $u \in \{x,y,z\} \cap B_k$;

(γ) if $\{x,y,z\} \cap B_k = \emptyset$ then $u = b_k$.

Since p_μ is a commutative Mal'tsev function, $|\{x,y,z\} \cap B_k| = 2$ is impossible and it is easy to see that $p\colon A^3 \to A$ is a commutative Mal'tsev function. We do not have to use semicentrality to check that p preserves ρ if $\mu \leqq \rho$ or $\rho \leqq \mu$; the trivial details will be omitted. Now let $\rho \in L$, $\rho \| \mu$, $x, x', y, z \in A$ and $x\rho x'$. We have to show that $p(x,y,z)\rho p(x',y,z)$. Suppose this is not the case. Since p preserves $\rho \circ \mu \in L$, we have $(p(x,y,z), p(x',y,z)) \in \rho \circ \mu = \rho \cup \cup \mu$ whence $p(x,y,z)\mu p(x',y,z)$. Therefore B_k in the definition of $p(x,y,z)$ and $p(x',y,z)$ is the same. If the same of (α), (β) and (γ) applies to both $p(x,y,z)$ and $p(x',y,z)$ then $p(x,y,z)\rho p(x',y,z)$. Moreover, if (α) applies to one of $p(x,y,z)$ and $p(x',y,z)$ then it applies to the other as well. Thus we may assume that (β) applies to $p(x,y,z)$ and (γ) applies to $p(x',y,z)$. Then $p(x,y,z) = x$, $p(x',y,z) = b_k$ and $x' \notin B_k$. From $b_k \mu x \rho x'$ and $\mu \circ \rho = \mu \cup \rho$ we conclude $(b_k, x') \in \rho$. Then we obtain $p(x',y,z) = b_k \rho x = p(x,y,z)$ from $x'\rho x$ and transitivity; this is a contradiction. Q.e.d.

Whatever it is evident the following lemma offers a comfortable way to exploit the permutability of L.

LEMMA 3. *Let $\mu, \rho \in L$, let B and C be distinct μ-blocks and suppose that $x\rho y$ for some $x \in B$, $y \in C$. Then*

$$\mathrm{SP}(\mu,\rho)\colon\ (\forall b \in B)(\exists c \in C)(b\rho c) \text{ and } (\forall c \in C)(\exists b \in B)(b\rho c).$$

(The notation SP stands for "shifting principle" and gives an economic way of referring to the lemma.)

The proof is a trivial application of the fact that $\mu \circ \rho = \rho \circ \mu$.

We say that an equivalence is of pattern $i_1 + i_2 + \cdots + i_t$ if it has t blocks and these blocks consists of $i_1, i_2, \ldots, i_t$ elements.

LEMMA 4. *If L has a member of pattern $j+1+1+\cdots+1$ where $1<j<n\leqq 8$ or $3+2+1+1+\cdots+1$ where $5\leqq n\leqq 8$ then we are done.*

PROOF. We will show that Lemma 2 is applicable. Assume that $\mu\in L$ is of pattern $j+1+\cdots+1$ and let B be the j-element block of μ. We claim that μ is semicentral. Indeed, if $(x,y)\in\mu\circ\rho=\rho\circ\mu$ but $(x,y)\notin\mu$ then, e.g., $x\notin B$ and $z\rho z\mu y$ holds for some $z\in A$. Since $[x]\mu$ is a singleton, $\mathrm{SP}(\mu,\rho)$ yields $(x,y)\in\rho$.

Now let μ be of pattern $3+2+1+\cdots+1$. Assume that μ is not semicentral. Let $B=\{a,b,c\}$ and $C=\{d,e\}$ be the nontrivial μ-blocks. We can consider a $\nu\in L$ and $x,y\in A$ with $(x,y)\in(\mu\circ\nu)\backslash(\mu\cup\nu)$. If $|\{x,y\}\cap\cap(B\cup C)|=1$, say $x\in B$, then $\mathrm{SP}(\mu,\nu)$ yields $(x,y)\in(B\cup\{y\})^2\subseteqq\nu$, a contradiction. Therefore $x\in B$ and $y\in C$ (or conversely). If $\nu|_C=1_C$ then $(x,y)\in(B\cup C)^2\subseteqq\nu$ by $\mathrm{SP}(\mu,\nu)$. Therefore $(d,e)\notin\nu$. Using $\mathrm{SP}(\mu,\nu)$ we have $B=\{z\in B: z\nu d\}\cup\{z\in B: z\nu e\}$ and we conclude that $\mu\cap\nu$ is of pattern $2+1+\cdots+1$. Therefore $\mu\cap\nu\in L$ is semicentral and Lemma 2 applies.

LEMMA 5. *If there are $\mu,\nu\in L$ such that*
$\mu<\nu$,
ν has exactly two blocks B and C,
$|B|>1$, $|C|>1$,
C is a block of μ and
there is a $b\in B$ with $[b]\mu=\{b\}$
then we are done.

PROOF. We intend to show that ν is semicentral. Assume that $\nu\circ\rho\neq\nu\cup\rho$ for some $\rho\in L$. Then there are $x,y\in A$ with $(x,y)\in\rho\backslash\nu$. By $\mathrm{SP}(\nu,\rho)$, there is a $c\in C$ with $b\rho c$. From $\mathrm{SP}(\mu,\rho)$ we conclude that $b\rho z$ holds for all $z\in C$. I.e., $C^2\subseteqq\rho$. Therefore $\mathrm{SP}(\nu,\rho)$ yields $\rho=1$, a contradiction. Q.e.d.

LEMMA 6. *Let $M_3=\{\omega,\alpha,\beta,\gamma,\iota\}$ be a diamond in L. Then every nontrivial block of ι/ω consists of four elements. The restriction of any of α/ω, β/ω and γ/ω to a four-element block of ι/ω has two two-element blocks. If ι/ω has only one nontrivial block (in particular, if $|A/\omega|<8$) then the interval $[\omega,\iota]$ of L coincides with M_3.*

PROOF. Since the ρ/ω (where $\omega\leqq\rho\in L$) permute, we can assume that $\omega=0$. Let B be a nontrivial ι/ω-block. Since M_3 is simple and the restriction map of M_3 to the equivalence lattice of B is a lattice homomorphism, $\{0_B,\alpha|_B,\beta|_B,\gamma|_B,1_B=\iota|_B\}$ is a diamond, too. It follows from Gumm [3, Lemma 3.2] and $|A/\omega|\leqq 8$ that $|B|=4$ and any of $\alpha|_B$, $\beta|_B$ and $\gamma|_B$ has two two-element blocks. We infer from Lemma 3 that beside $\alpha|_B$, $\beta|_B$ and $\gamma|_B$ no nontrivial equivalence on B permute with $\alpha|_B$, $\beta|_B$ and $\gamma|_B$ simultaneously. Thus $[0,\iota]=M_3$, provided B is the only nontrivial block of ι/ω. Q.e.d.

In virtue of Lemma 1 we have to prove our theorem only for those cases when L includes a diamond $M_3=\{\omega,\alpha,\beta,\gamma,\iota\}$. L can include more than one diamond but $M_3=\{\omega,\alpha,\beta,\gamma,\iota\}$ will always denote a fixed diamond for

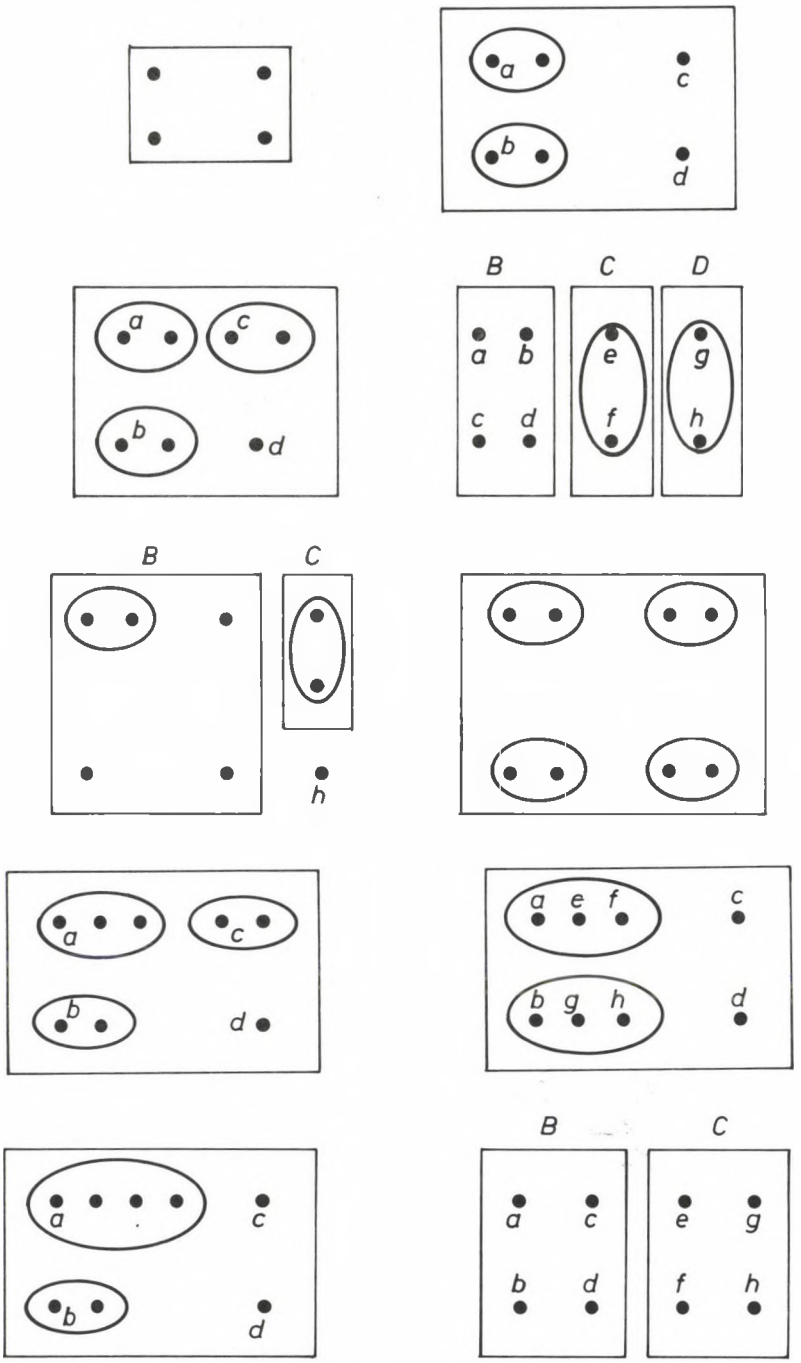
which ω is minimal. It is well-known in the theory of modular lattices that if a modular lattice M has a diamond whose bottom is $x \in M$ then there is an interval $[x, y]$ of length two which includes a diamond. (Having no simple reference at hand we refer to the far more general Freese [1, Thm. 1.7].) Therefore we always assume that our fixed diamond M_3 with minimal ω also satisfies $\omega -\!\!< \alpha -\!\!< \iota$, $\omega -\!\!< \beta -\!\!< \iota$ and $\omega -\!\!< \gamma -\!\!< \iota$. By Lemma 6 we do not have too many possibilities for M_3. Moreover, if Lemma 4 or Lemma 5 applies for ω and/or ι then we are done. Now it is easy to check that we are left with ten cases only; they are depicted on Figs. 1–10. On these figures, the nontrivial ι-blocks are denoted by rectangles while the nontrivial ω-blocks, if there is any, are encircled. When some or all of the elements of A are labelled, we always assume that $(a, d), (b, c) \in \alpha$, $(b, d), (a, c) \in \beta$ and $(c, d), (a, b) \in \gamma$; this convention generally determines α, β and γ in virtue of Lemma 6. Sometimes ι-blocks are labelled with capital letters.

In Case 1 (cf. Fig. 1) we can equip A with an Abelian group structure so that A be of exponent two and $\mathrm{Con}(A) = L$. Then $p(x, y, z) = x + y + z$ is a commutative Mal'tsev function compatible with L.

In Cases 2, 3, 7, 8 and 9 we are going to show that for any other diamond $\{\omega', \alpha', \beta', \gamma', \iota'\}$ in L we have $\omega \leqq \omega'$. (Then Lemma 1 is applicable with $\mu = \omega$.) Suppose this is not the case, i.e., $\omega || \omega'$. We intend to show that ω' must have less than four blocks, which contradicts Lemma 6. Take an $(x, y) \in \omega' \backslash \omega$. Using $\mathrm{SP}(\gamma, \omega')$ or $\mathrm{SP}(\beta, \omega')$ we may assume that $x = d$. If $y \in [a]\omega$ then $\mathrm{SP}(\omega, \omega')$ yields $([a]\omega \cup \{d\})^2 \subseteqq \omega'$ and, by using $\mathrm{SP}(\beta, \omega')$, we can see that ω' has at most $|[c]\omega| \leqq 2$ further blocks beside $[a]\omega'$. Similarly, if $y \in [b]\omega$ then $\mathrm{SP}(\omega, \omega')$ yields $([b]\omega \cup \{d\})^2 \subseteqq \omega'$ and, by $\mathrm{SP}(\gamma, \omega')$, ω' has at most $|[c]\omega| + 1 \leqq 3$ blocks. Now suppose $x \in [c]\omega$. Then, by $\mathrm{SP}(\omega, \omega')$, $\{d\} \cup [c]\omega \subseteqq [d]\omega'$. If $|[a]\omega| < 3$ or $|[b]\omega| < 3$ then, by $\mathrm{SP}(\beta, \omega')$, ω' has at most three blocks. Therefore ω' may have four blocks only in Case 8 and, apart from labelling, these blocks are $\{a, b\}$, $\{e, g\}$, $\{f, h\}$ and $\{c, d\}$. By Lemma 6, $\rho = abeg; fhcd \in [\omega', \iota'] \subseteq L$. (Here and often in the sequel an equivalence relation is denoted by the list of its nontrivial blocks separated by semicolons.) Hence $\mathrm{SP}(\rho, \omega)$ leads to a contradiction.

To settle Case 4, assume that ι is not semicentral. Then there is a $\rho \in \in L\backslash\{1\}$ such that $(x, y) \in \rho \backslash \iota$. If $\rho \subseteqq B^2 \cup (C \cup D)^2$ then Lemma 2 applies for $\iota + \rho = abcd; efgh$, which is semicentral. Indeed, if we had, e.g., $(a, e) \in \nu \backslash (\iota + + \rho)$ for some $\nu \in L\backslash\{1\}$ then $\mathrm{SP}(\omega, \nu)$ would give $[a]\nu \supseteqq \{a, e, f\}$, $\mathrm{SP}(\iota, \nu)$ would yield $[a]\nu \supseteqq B \cup C$ and $\mathrm{SP}(\nu, \rho)$ would lead to a contradiction since $[g]\rho \cap C \neq \emptyset$ and $[h]\rho \cap C \neq \emptyset$ by $\mathrm{SP}(\omega, \rho)$. Therefore $(x, y) = (a, e) \in \rho$ can be assumed. Then $[a]\rho \supseteqq B \cup C$ like in case of ν before. Hence $[a]\rho = B \cup C$ as otherwise $\mathrm{SP}(\iota, \rho)$ would lead to $\rho = 1$. Now either Lemma 4 applies for ρ or Lemma 5 applies for $\omega < \rho$.

The treatment for Case 5 starts with assuming that ι is not semicentral. Then $\rho \circ \iota \neq \rho \cup \iota$ for some $\rho \in L\backslash\{0, 1\}$. If $[h]\rho = \{h\}$ then $B \cup C$ is the only nontrivial block of $\rho + \iota$ and Lemma 4 applies. Observe that $[h]\rho \cap$



Figs. 1–10

$\cap B \neq \emptyset$ implies $(B \cup \{h\})^2 \subseteq \rho$ and $[h]\rho \cap C \neq \emptyset$ implies $(C \cup \{h\})^2 \subseteq \rho$ by $\mathrm{SP}(\iota, \rho)$, but only one of these two possibilities can occur as $\rho \neq 1$. Therefore if $[h]\rho \neq \{h\}$ then Lemma 5 applies for ι and $\iota + \rho$.

In Case 6 we may assume by Lemma 1 that there exists another diamond $M_3' = \{\omega', \alpha', \beta', \gamma', \iota'\}$ with $\omega \not\leqq \omega'$. We choose this M_3' so that ω' be minimal. Like in case of M_3 we may assume that $\omega' -\!\!< \alpha' -\!\!< \iota'$, $\omega' -\!\!< \beta' -\!\!< \iota'$ and $\omega' -\!\!< \gamma' -\!\!< \iota'$. Since $\omega' || \omega$ and the previous cases have been handled, we may suppose that ω' is also of pattern $2+2+2+2$. As $\omega' || \omega$, they can have 0, 1 or 2 blocks in common. However, if they had exactly one block in common then Lemma 4 would apply to $\omega' \cap \omega$; if they had two blocks, say $\{a, e\}$ and $\{b, f\}$, in common then $\mathrm{SP}(\alpha, \omega')$ would lead to a contradiction. Therefore, by $\mathrm{SP}(\omega, \omega')$, we may assume that the situation is as depicted on Figure 11, where the horizontal lines indicate ω'. Since the role of α', β' and γ' is symmetric, we assume that $\alpha' = abcd; efgh$, $\beta' = abef; cdgh$ and $\gamma' = abgh; cdef$. Let $\mathbf{Z}_2 = \{0, 1\}$ denote the two-element Abelian group. We consider A the (support of) $\mathbf{Z}_2^3$ as indicated on Figure 11. Since $\mathrm{Con}(\mathbf{Z}_2^3)$ admits a commutative Mal'tsev function $p(x, y, z) = x + y + z$, it suffices to show that $L \subseteq \mathrm{Con}(\mathbf{Z}_2^3)$. If $0 < \rho \leqq \omega$ for $\rho \in L$ then $\rho = \omega$ by $\mathrm{SP}(\alpha', \rho)$. I.e., ω is an atom in L. So is ω', for the role of M_3 and M_3' is symmetric. If $\rho \in L$ is in $[\omega, \iota] = [\omega, 1]$ or $[\omega', \iota'] = [\omega', 1]$ then $\rho \in \mathrm{Con}(\mathbf{Z}_2^3)$ by Lemma 6 and $M_3, M_3 \subseteq \mathrm{Con}(\mathbf{Z}_2^3)$. Suppose $\rho \in L \backslash \{0\}$ but $\omega \not\leqq \rho$, $\omega' \not\leqq \rho$. Then $\rho \cap \omega = \rho \cap \omega' = 0$. If $\rho \leqq \omega + \omega'$ then a standard argument with the height function of L yields that $\{0, \omega, \omega', \rho, \omega + \omega'\}$ is a diamond, which contradicts the minimality of ω. Hence $\rho \not\leqq \omega + \omega'$, whence $x \rho y$ holds for some $x \in \{a, b, e, f\}$ and $y \in \{c, d, g, h\}$. We can suppose $x = a$ by $\mathrm{SP}(\omega, \rho)$ and $\mathrm{SP}(\omega', \rho)$. Since the possibilities $a \rho d$, $a \rho c$, $a \rho g$ and $a \rho h$ are quite analogous, we detail the case $a \rho d$ only. Then using $\mathrm{SP}(\omega, \rho)$ and $\mathrm{SP}(\omega', \rho)$ we derive $\rho \supseteq ad; bc; fg; eh$. If $\rho = ad; bc; fg; eh$ then $\rho \in \mathrm{Con}(\mathbf{Z}_2^3)$. So suppose $\rho \supset ad; bc; fg; eh$. Since $\rho \cap \omega = \rho \cap \omega' = 0$, it follows either $a \rho f$ or $b \rho e$. By $\mathrm{SP}(\omega, \rho)$ both hold. Hence $\rho \supseteq adfg; bceh$. Since $\rho \neq 1$, $\rho = adfg; bceh \in \mathrm{Con}(\mathbf{Z}_2^3)$.

In Case 10, the restriction map to any block of ι is injective, for it does not collapse $\omega = 0$ and ι. Therefore $\alpha = ad; bc; eh; fg$, $\beta = bd; ac; eg; fh$ and $\gamma = cd; ab; ef; gh$ can be assumed. We consider A as $\mathbf{Z}_2^3$ exactly the same way as before. We intend to show $L \subseteq \mathrm{Con}(\mathbf{Z}_2^3)$. Evidently, $M_3 = \{0, \alpha, \beta, \gamma, \iota\} \subseteq \mathrm{Con}(\mathbf{Z}_2^3)$. To show $[0, \iota] = M_3$ assume that $0 < \rho < \iota$, $\rho \in L \backslash M_3$. Applying Lemma 6 to $\{\mu|_B : \mu \in [0, \iota]\}$ and $\{\mu|_C : \mu \in [0, \iota]\}$ we derive that the restriction of ρ to either block of ι coincides with the restriction of a member of M_3. E.g., $\rho|_B = \alpha|_B$ but $\rho|_C \neq \alpha|_C$. Then $\rho|_C \neq \iota|_C$ implies $0 < \rho \cap \alpha < \alpha$ while $\rho|_C = \iota|_C$ yields $\alpha < \rho < \iota$, both contradicting $0 -\!\!< \alpha -\!\!< \iota$. Having seen that $[0, \iota] \subseteq \mathrm{Con}(\mathbf{Z}_2^3)$ let us assume that $\rho \not\leqq \iota$, $\rho \in L \backslash \{1\}$. Then, e.g., $a \rho e$. Now $\mathrm{SP}(\gamma, \rho)$ gives $b \rho f$ and $\mathrm{SP}(\alpha, \rho)$ gives $ae; bf; cg; dh \subseteq \rho$. If we have equality then $\rho \in \mathrm{Con}(\mathbf{Z}_2^3)$. If $ae; bf; cg; dh \subset \rho$ then $\rho \cap \alpha \neq 0$ or $\rho \cap \beta \neq 0$ or $\rho \cap \gamma \neq 0$. E.g., suppose $\rho \cap \alpha \neq 0$. As α is an atom, $\rho \geqq \alpha$. Hence

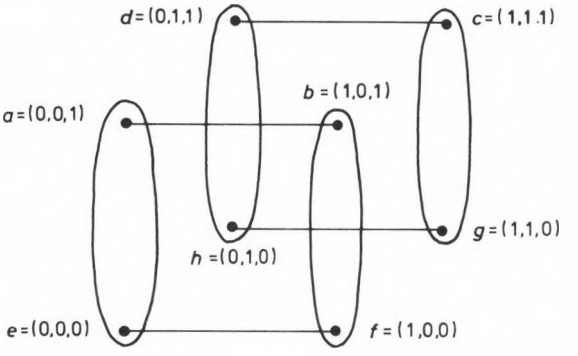


Fig. 11

$\rho \geqq aedh; bfcg$. I.e., $\rho = 1$ or $\rho = aedh; bfcg$, whence $\rho \in \mathrm{Con}(\mathbf{Z}_2^3)$. Q.e.d.

REFERENCES

[1] FREESE, R., The variety of modular lattices is not generated by its finite members, *Trans. Amer. Math. Soc.* **255** (1979), 277–300. *MR* **81g**:06003

[2] GRÄTZER, G., *General lattice theory*, Pure and Applied Mathematics, 75, Academic Press, Inc., New York – London, 1978, *MR* **80c**:06001b

[3] GUMM, H.-P., Is there a Mal'cev theory for single algebras?, *Algebra Universalis* **8** (1978), 320–329. *MR* **57** #12343

[4] GUMM, H.-P., Algebras in permutable varieties: geometrical properties of affine algebras, *Algebra Universalis* **9** (1979), 8–34. *MR* **80d**:08010

[5] MAL'CEV, A. I., On the general theory of algebraic systems, *Mat. Sbornik N. S.* **35** (**77**) (1954), 3–20 (in Russian). *MR* **16**–440

[6] PIXLEY, A. F., Completeness in arithmetical algebras, *Algebra Universalis* **2** (1972), 179–196. *MR* **48** #208

(Received March 12, 1990)

DEPARTMENT OF ALGEBRA AND GEOMETRY
PALACKY UNIVERSITY OLOMOUC
TOMKOVA 38
779 00 OLOMOUC
CZECH REPUBLIC

JÓZSEF ATTILA TUDOMÁNYEGYETEM
BOLYAI INTÉZETE
ARADI VÉRTANUK TERE 1
H–6720 SZEGED
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 349–358

ON k-CRITICAL GRAPHS WITH MANY EDGES AND NO SHORT CYCLES

H. L. ABBOTT and B. ZHOU

Abstract

Denote by $f_{k,l}(n)$ the largest integer for which there exists a k-critical graph of order n having girth at least $l+1$ and $f_{k,l}(n)$ edges. We prove that for $l=4$ or 5 and for each $k \geqq 4$ there exists a positive constant c_k such that $f_{k,l}(n) > c_k n^{3/2}$ for all sufficiently large n.

Let k be an integer, $k \geqq 3$. A graph G is said to be k-critical if it has chromatic number k, but every proper subgraph of G is $(k-1)$-colorable. It is an old result of König [10] that the only 3-critical graphs are the cycles of odd length. G. A. Dirac [4] proved that if $k \geqq 4$, $n \geqq k$ and $n \neq k+1$ then there exists a k-critical graph of order n. For $n = k$ the only such graph is, of course, the complete graph of order k. We suppose in what follows that $k \geqq 4$ and $n \geqq k+2$. P. Erdős raised the following question: What is the largest integer $f_k(n)$ for which there exists a k-critical graph of order n with $f_k(n)$ edges? Dirac [5] observed that if C_1 and C_2 are cycles of length $2t+1$ and if each vertex of C_1 is joined by an edge to each vertex of C_2 the resulting graph is 6-critical. Thus, if $n = 4t+2$,

$$f_6(n) \geqq n^2/4 + n. \tag{1}$$

Dirac also remarked that if G is a k-critical graph of order n, the graph obtained by joining a new vertex to each vertex of G is a $k+1$-critical graph of order $n+1$ and thus

$$f_{k+1}(n+1) \geqq f_k(n) + n. \tag{2}$$

It follows from (1) and (2) that for each $k \geqq 6$ there exists a positive constant α_k such that for infinitely many integers n

$$f_k(n) \geqq \alpha_k n^2. \tag{3}$$

The above constructions of Dirac yield no information in the cases $k = 4$ or 5. That (3) holds for infinitely many n in each of these cases was established by B. Zeidl [20] and B. Toft [15]. In fact, Toft showed that if $k = 3q+r$,

1980 *Mathematics Subject Classification* (1985 Revision). Primary 05C15; Secondary 05C35.

Key words and phrases. Critical graphs, short cycles.

Akadémiai Kiadó, Budapest

$0 \leqq r \leqq 2$, then for infinitely many n

$$f_k(n) > \begin{cases} \dfrac{q-1}{2q}n^2 & \text{if } r=0 \\ \dfrac{7q-6}{14q+2}n^2 & \text{if } r=1 \\ \dfrac{23q-15}{46q+16}n^2 & \text{if } r=2. \end{cases} \tag{4}$$

A proof of the first inequality had been given earlier by Dirac and Erdös [6]. It follows from the classical theorem of Turán [18] that

$$f_k(n) < \frac{k-2}{2k-2}n^2. \tag{5}$$

Note that if β_k is taken to be the supremum of the numbers α_k for which (3) holds for infinitely many n then from (4) and (5) it follows that $\lim_{k\to\infty} \beta_k = \frac{1}{2}$. However, β_k has not been determined for any value of k.

We propose in this paper to investigate a variant of the problem described above. It is a classical result in graph coloring theory that for each pair k, l, $k \geqq 4$, $l \geqq 3$, there exist k-critical graphs with no cycles of length at most l (that is, whose girth is at least $l+1$). In fact, there exists a least integer $N(k,l)$ such that if $n \geqq N(k,l)$ there exists such a graph of order n. For $n \geqq N(k,l)$ denote by $f_{k,l}(n)$ the largest integer for which there exists a k-critical graph of order n whose girth is at least $l+1$ and which has $f_{k,l}(n)$ edges.

Consider first the case $l=3$. The following construction is given in [15]. It is one of the constructions used in establishing (6). Let t be an odd integer, $t \geqq 5$. Let A_1 and A_2 denote the color classes of the complete bipartite graph $K_{t,t}$. Let C_1 and C_2 be cycles of length t and set up matchings from C_1 to A_1 and from C_2 to A_2. The resulting graph clearly has no triangles and it is shown in [15] that it is 4-critical. It follows that if $n=4t$, t odd, then

$$f_{4,3}(n) \geqq \frac{n^2}{16} + n. \tag{6}$$

By the well-known construction of Mycielski [12] we get

$$f_{k+1,3}(2n+1) \geqq 2f_{k,3}(n) + n \tag{7}$$

and it follows from (6) and (7) that for $k \geqq 4$

$$f_{k,3}(n) \geqq \frac{n^2}{2^k} \tag{8}$$

holds for infinitely many n. One may in fact show that for each $k \geqq 4$ there exists a positive constant γ_k such that

$$f_{k,3}(n) \geqq \gamma_k n^2 \tag{9}$$

for all sufficiently large n. It would be of interest to decide whether $f_{k,3}(n) > > \gamma n^2$ holds for some positive constant γ which does not depend on k.

The nature of the problem changes substantially for $l \geqq 4$. Denote by $g(n)$ the largest integer for which there exists a graph of order n having girth at least 5 and $g(n)$ edges. It is known that, as $n \to \infty$,

$$g(n) = \left(\frac{1}{2} + o(1)\right) n^{3/2}. \tag{10}$$

This result is due to Erdős, Rényi and Sós [8], Erdős and Rényi [7] and Brown [2]. Since $f_{k,l}(n) \leqq g(n)$ we get

$$f_{k,l}(n) < cn^{3/2} \tag{11}$$

for all $c > \frac{1}{2}$ and all sufficiently large n. The graphs which show that equality holds in (10) are bipartite and thus give no information concerning lower bounds for $f_{k,l}(n)$. There are many constructions (explicit and random) of graphs with prescribed chromatic number and girth. See [17], Section 2, for an account of some of these and for references to the literature. However, the graphs constructed in many of these papers have not been proven to be color-critical, and for those constructions for which this has been done the graphs in question are very sparse and give lower bounds for $f_{k,l}(n)$ which grow only linearly with n. For example, the well-known construction of Tutte [19] gives only $f_{4,l}(n) \geqq (2 + o(1))n$ for $l = 4$ or 5. Tutte's construction gives k-chromatic graphs with girth 6. However, these are not known to be k-critical except in the case $k = 4$.

We now state our main result.

THEOREM. *For $l = 4$ or 5 and for each $k \geqq 4$ there exists a positive constant $c = c(k, l)$ such that for all $n \geqq n(k, l)$*

$$f_{k,l}(n) > cn^{3/2}. \tag{12}$$

PROOF. The construction establishing (12) is complicated. We show first that it holds for $k = 4$ and $l = 5$. The graph constructed by Toft which was described earlier in connection with the proof of (6) has a large number 4 and 5-cycles, all of which involve edges from the bipartite portion of the graph. We wish to destroy these cycles. This may be achieved via splitting operations similar to those described in [3], [11], [14], [15] and [16]. However, in doing this the number of new vertices added is very large and the number of new edges added is of the same order of magnitude so that the lower

bound obtained for $f_{4,l}(n)$ grows only linearly with n. The novelty of our argument involves combining these splitting operations with the geometric constructions of [2] and [8] that are used in establishing (10).

Let G be a 4-critical graph of girth at least 6 with a vertices and b edges. Let uv be an edge of G; we refer to uv as the special edge. Let q be a prime, $q \geqq 5$, and let $G_1, G_2, \ldots, G_{q^2-1}$ be copies of G. Denote the special edge of G_i by $u_i v_i$. Delete the special edges, identify v_i with u_{i+1} for $i = 1, 2, \ldots, q^2 - 2$ and relabel v_{q^2-1} as u_{q^2}. Denote the resulting graph by G^*. Note that G^* has $(a-1)q^2 - a + 2$ vertices and $(b-1)q^2 - b + 2$ edges. G^* is 3-colorable but in any 3-coloring the special vertices $u_1, u_2, \ldots, u_{q^2}$ are assigned the same color.

Let Z_q denote the finite field with q elements and consider the affine plane geometry P over Z_q. For $i = 1, 2, \ldots, q$ and $j = 1, 2, \ldots, q$ let

$$L_{i,j} = \{(x, y) : x, y \in Z_q,\ y = ix + j\}$$

and for $j = 1, 2, \ldots, q$ let

$$L_{q+1,j} = \{(j, y) : y \in Z_q\}.$$

Note that we are denoting the zero element of Z_q by q instead of 0. Less formally, $L_{i,j}$ is the set of points on the line whose equation is $y = ix + j$, if $i \neq q+1$, and $L_{q+1,j}$ is the set of points on the "vertical" line $x = j$. This gives $q+1$ partitions $P_1, P_2, \ldots, P_{q+1}$ of the points of P where P_i is given by

$$P_i = L_{i,1} \cup L_{i,2} \cup \ldots \cup L_{i,q}.$$

We refer to $L_{i,j}$ as the jth part of P_i. Note that if $i_1 \neq i_2$ then each part of P_{i_1} intersects each part of P_{i_2} in exactly one element of P. It will be convenient to denote the points of P by $z_1, z_2, \ldots, z_{q^2}$ and to suppose that the labelling is such that the first part of P_1 is $\{z_1, z_2, \ldots, z_q\}$; equivalently, that $L_{1,1} = \{z_1, z_2, \ldots, z_q\}$.

Let $C_1, C_2, D_1, D_2, \ldots, D_q$ be cycles of length q. Let $A = \{a_1, a_2, \ldots, a_q\}$ and $B = \{b_1, b_2, \ldots, b_q\}$ be sets of size q. It is understood that the sets $A, B, S_1, S_2, \ldots, S_q$ and the vertex sets of $C_1, C_2, D_1, D_2, \ldots, D_q$ are pairwise disjoint and are disjoint from the vertex set of G^* and the points of the plane P.

Let H_q be the graph whose vertex set consists of the following points (the reader will find it useful to examine Fig. 1):

(i) The vertex sets of $C_1, C_2, D_1, D_2, \ldots, D_q$.
(ii) The elements of A, B and $S_1, S_2, \ldots, S_q$.
(iii) The points of the plane P.
(iv) The vertex set of G^*.

The edges of H_q are as follows:

(a) The edges of the cycles $C_1, C_2, D_1, D_2, \ldots, D_q$.

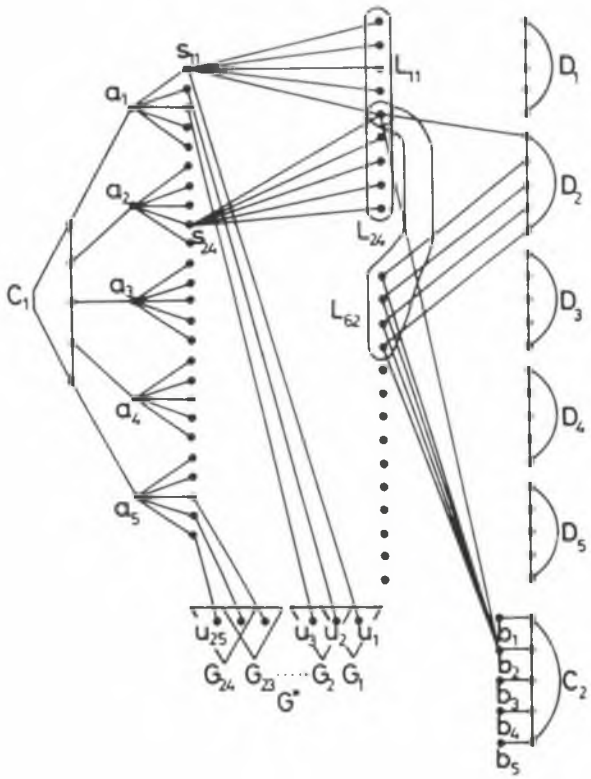


Fig. 1
The graph H_5. Not all edges are shown

(b) A matching from C_1 to A.

(c) A matching from C_2 to B.

(d) The edges $a_i s_{i,j}$, $i = 1, 2, \ldots, q$; $j = 1, 2, \ldots, q$.

(e) All edges joining b_j to each point of the jth part of P_{q+1}, $j = 1, 2, \ldots, q$.

(f) A matching from D_j to the jth part of P_{q+1}, $j = 1, 2, \ldots, q$.

(g) All edges of the form $s_{i,j} z_m$ where z_m is in the jth part of P_j. Note that there are q^3 such edges.

(h) A matching from $\{u_1, u_2, \ldots, u_{q^2}\}$ to $\cup S_i$.

(i) The edges of G^*.

One easily verifies that the number of vertices of H_q is $(a+2)q^2 + 4q - a + 2$ and the number of edges is $q^3 + (b+3)q^2 + 4q - b + 2$.

We verify that H_q is 4-chromatic. Suppose that H_q has a 3-coloring in colors red, blue and green, say. Then $u_1, u_2, \ldots, u_{q^2}$ must be assigned the same color, say red. This implies, because of (h), that each vertex of

$\cup S_i$ must be colored blue or green. If, for each i, S_i has both blue and green vertices then all vertices of A must be colored red, because of (d). This implies, via (b), that the vertices of C_1 must be colored blue or green, contradicting the fact that C_1 is a cycle of odd length. Hence for some i the vertices of S_i are monochromatic; suppose they are blue. Since each vertex of P is joined to some vertex of S_i, no vertex of P is colored blue. Because of (f), each part of P_{q+1} must contain both red and green vertices; otherwise the vertices of some odd cycle D_j cannot be properly colored. This implies, by (e), that all vertices of B must be colored blue. But then, because of (c), we cannot properly color C_2. It follows that H_q is not 3-colorable. It is clearly 4-colorable and thus 4-chromatic.

We do not know whether H_q is 4-critical. We show, however, that every 4-critical subgraph of H_q must contain all of the edges of type (g) and thus at least q^3 edges. Delete an edge e of type (g). We shall show that $H_q - e$ is 3-colorable. There is no loss of generality in supposing that $e = s_{1,1}z_1$. Recall that the labelling is such that $s_{1,1}$ is joined to $z_1, z_2, \ldots, z_q$. There is also no loss of generality in supposing that z_1 is adjacent to $s_{2,1}, s_{3,1}, \ldots, s_{q,1}$. 3-color G^* in colors red, blue and green so that the special vertices $u_1, u_2, \ldots, u_{q^2}$ are green. Color the vertices of P and $\cup S_i$ as follows:

red	blue	green
z_1	$z_2, z_3, \ldots, z_q$	$z_{q+1}, z_{q+2}, \ldots, z_{q^2}$
$s_{1,j},\ j = 1, 2, \ldots, q$	$s_{2,1}, s_{3,1}, \ldots, s_{q,1}$	
$s_{i,j},\ i = 1, 2, \ldots, q;\ j = 2, 3, \ldots, q$		

There are obviously no green edges (at this stage). There are also no red edges, for if $s_{i,j}z_1$, $j \neq 1$, were a red edge, we would have $i = 1$ so that z_1 is in the jth part of P_1, contrary to the fact that z_1 is in the first part of P_1. Suppose there were a blue edge, say $s_{i,1}z_m$, $2 \leqq i \leqq q$, $2 \leqq m \leqq q$. Then z_m is in $L_{i,1}$ and also in $L_{1,1}$ so that $L_{i,1} \cap L_{1,1} = \{z_m\}$. However, by (g), this implies that z_1 and z_m are both in the first parts of P_1 and P_i and this is not so. Hence there are no blue edges.

We now extend the coloring to the rest of the graph. Color a_1 blue and a_j green for $j = 2, 3, \ldots, q$. Color the vertex of C_1 incident with a_1 green and the remaining vertices alternately red and blue. For each j, there is one vertex in the jth part $L_{q+1,j}$ of P_{q+1} colored red or blue and the remaining vertices of $L_{q+1,j}$ are colored green. Color the vertex of D_j which is adjacent to the red or blue vertex of $L_{q+1,j}$ green and the remaining vertices alternately red and blue. Color b_j red if it is not adjacent to z_1. In this case b_j will have one blue neighbor in P and $q - 1$ green neighbors. If b_j is adjacent to z_1, it cannot be adjacent to any of $z_2, z_3, \ldots, z_q$, so that we may then color b_j blue. Thus precisely one vertex b of B is colored blue and all others are red. We may then color the vertex of C_2 adjacent to b red and the remaining vertices of C_2 alternately blue and green. This gives the desired 3-coloring of $H_q - e$.

Let H_q' denote a 4-critical subgraph of H_q of maximal order. It is straightforward to check that H_q, and hence also H_q', has no cycles of length at most 4. In fact, it is clear that if there is such a cycle, it must have length 4 and must involve one vertex from $\cup S_i$, a vertex b_m of B and two vertices z_i and z_j of P. This implies that z_i and z_j are in the mth part of P_{q+1} and since this part intersects each part of P_t, for any $t < q+1$ exactly once, z_i and z_j cannot have a common neighbor in $\cup S_i$. Thus there are no cycles of length 4 either. H_q does have cycles of length 5 and thus H_q' may also have 5-cycles. However, it is clear that any such 5-cycle must have an edge e which is not of type (g), (h) or (i). Recall the well-known construction of Hajós [9]. Let Γ_1 and Γ_2 be k-critical graphs with girth at least $l+1$ and let a_1b_1 and a_2b_2 be edges of Γ_1 and Γ_2. Delete a_1b_1 and a_2b_2, identify a_1 with a_2 and join b_1 and b_2 by a new edge. The resulting graph is k-critical and also has no cycles of length at most l. Let e be an edge of H_q' not of type (g), (h) or (i) and which lies on a 5-cycle. Apply the Hajós construction to H_q' and a copy of G with e playing the role of one of the special edges. The 5-cycle is then destroyed. It therefore follows that by applying the Hajós construction at most $4q^2+4q$ times, all of the 5-cycles of H_q' will be destroyed. Denote the resulting graph by H_q^*. If H_q^* has n_q vertices and m_q edges, then

$$2q^2 \leqq n_q \leqq (5a+2)q^2+(4a+4)q \leqq 6aq^2 \tag{13}$$

and

$$q^3 \leqq m_q \leqq q^3+(5b+3)q^2+(4b+4)q. \tag{14}$$

Furthermore we have

$$f_{4,5}(n_q) \geqq m_q. \tag{15}$$

Let $p=h(q)$ be the least prime such that $2p^2 > n_q$. Then, from (13), we get

$$2q^2 \leqq n_q < 2p^2 \leqq n_p \leqq 6ap^2. \tag{16}$$

By the theorem of Chebyshev (see [13], p. 131–136) asserting that there is a prime between x and $2x$ for all $x \geqq 2$ we have $p < (2n_q)^{1/2}$ so that $p^2 < 2n_q$. It then follows from (16) that

$$n_q < n_p < 12an_q. \tag{17}$$

Let $q_1 = 5$ and for $i \geqq 1$ let $q_{i+1} = h(q_i)$. We then have, from (15),

$$f_{4,5}(n_{q_i}) \geqq m_{q_i}. \tag{18}$$

For $N(4,5) \leqq r \leqq n_{q_{i+1}} - n_{q_i} + N(4,5) - 1$, let Γ_r be a 4-critical graph with r vertices, s edges and girth at least 6. Apply the Hajós construction to $H_{q_i}^*$

and Γ_r. This gives a 4-critical of girth at least 6 having $n_{q_i} + r - 1$ vertices and $m_{q_i} + s - 1$ edges. It follows that

$$f_{4,5}(n_{q_i} + r - 1) \geqq m_{q_i} + s - 1. \tag{19}$$

Let n be a large integer. Determine i by $n_{q_i} + N(4,5) - 1 \leqq n \leqq n_{q_{i+1}} + N(4,5) - 2$. Then $n = n_{q_i} + r - 1$ for some r in $[N(4,5), n_{q_{i+1}} - n_{q_i} + N(4,5) - 1]$. We then get

$$\begin{aligned} f_{4,5}(n) &= f_{4,5}(n_{q_i} + r - 1) \\ &\geqq m_{q_i} + s - 1, \quad \text{by (19)} \\ &> q_i^3, \quad \text{by (14)} \\ &\geqq \left(\frac{1}{6a} n_{q_i}\right)^{3/2}, \quad \text{by (13)} \\ &\geqq \left(\frac{1}{72a^2} n_{q_i}\right)^{3/2}, \quad \text{by (17)} \\ &\geqq \left(\frac{1}{72a^2}(n - N(4,5) + 2)\right)^{3/2}, \quad \text{by (19)} \\ &> cn^{3/2} \end{aligned}$$

where c depends only on a.

We have thus proved (12) for $k = 4$. We now proceed by induction on k. Suppose $k \geqq 4$ and that (12) has been established for k; that is, suppose that we have established the existence of an integer $n(k)$ and a constant c_k such that

$$f_{k,5}(n) > c_k n^{3/2}$$

for all $n \geqq n(k)$. Let Γ be a $(k+1)$-critical graph with t vertices and s edges and having no cycles of length at most 5. Let Γ^* be the graph obtained from $n - 1$ copies of Γ in the way in which G^* was constructed from G. Let the special vertices of Γ^* be $u_1, u_2, \ldots, u_n$. Let H be the graph constructed by setting up a matching from the vertices $u_1, u_2, \ldots, u_n$ of Γ^* and the vertex set of a k-critical graph of order n and girth 6 with $f_{k,5}(n)$ edges. It is straightforward to check that H is $(k+1)$-critical and has girth 6. H has $t(n-1) + 2$ vertices and $f_{k,5}(n) + s(n-1) + 1$ edges.

Thus

$$f_{k+1,5}(t(n-1) + 2) \geqq f_{k,5}(n) + s(n-1) + 1.$$

Here t depends only on $k+1$. We could for example take $t = N(k+1, 5)$. It follows that

$$f_{k+1,5}(n) \geqq \frac{c_k}{t^{3/2}} n^{3/2} = \frac{c_k}{(N(k+1,5))^{3/2}} n^{3/2}.$$

Thus (12) holds for $k+1$ for infinitely many integers n. One may now appeal to the Hajós construction again in order to ensure that (12) holds for $k+1$ for all sufficiently large n. We suppress the details of this part of the argument.

We remark that the constants obtained in the above proof are very small. It would be of interest to improve them. We do not know what is the order of magnitude of $f_{k,l}(n)$ for $l \geqq 6$. We hope to return to this question in a later paper.

REFERENCES

[1] Bollobás, B., *Extremal graph theory*, London Mathematical Society Monographs, 11, Academic Press, London–New York, 1978. *MR* **80a**:05120

[2] Brown, W. G., On graphs that do not contain a Thomsen graph, *Canad. Math. Bull.* **9** (1966), 281–285. *MR* **34** #81

[3] Brown, W. G. and Moon, J. W., Sur les ensembles de sommets indépendents dans les graphes chromatiques minimaux, *Canad. J. Math.* **21** (1969), 274–278. *MR* **38** #4358

[4] Dirac, G. A., Circuits in critical graphs, *Monatsh. Math.* **59** (1955), 178–187. *MR* **17**–289

[5] Dirac, G. A., A property of 4-chromatic graphs and some remarks on critical graphs, *J. London Math. Soc.* **27** (1952), 85–92. *MR* **13**–572

[6] Erdős, P., Problems and results in chromatic graph theory, *Proof Techniques in Graph Theory* (Proc. Second Ann Arbor Graph Theory Conf., Ann Arbor, Mich., 1968), Academic Press, New York, 1969, 27–35. *MR* **40** #5494

[7] Erdős, P. and Rényi, A., On a problem in the theory of graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **7** (1962), 623–641. *MR* **33** #1246

[8] Erdős, P., Rényi, A. and Sós, V. T., On a problem of graph theory, *Studia Sci. Math. Hungar.* **1** (1966), 215–235. *MR* **36** #6310

[9] Hajós, G., Über ein Konstruktion nicht n-färbbarer Graphen, *Wiss. Z. Martin-Luther-Univ. Halle-Wittenberg Math.-Natur. Reihe* **10** (1961), 116–117.

[10] König, D., *Theorie der endlichen und unendlichen Graphen. Kombinatorische Topologie der Streckenkomplexe*, Chelsea, New York, 1950. *MR* **12**–195

[11] Lovász, L., Independent sets in critical chromatic graphs, *Studia Sci. Math. Hungar.* **8** (1973), 165–168. *MR* **48** #8289

[12] Mycielski, J., Sur le coloriage des graphes, *Colloq. Math.* **3** (1955), 161–162. *MR* **16**–1044

[13] Sierpiński, W., *Elementary theory of numbers*, Monografie Matematyczne, Tom 42, Państwowe Wydawnictwo Naukowe, Warszawa, 1964. *MR* **31** #116

[14] Simonovits, M., On colour-critical graphs, *Studia Sci. Math. Hungar.* **7** (1972), 67–81. *MR* **47** #4841

[15] Toft, B., On the maximal number of edges of critical k-chromatic graphs, *Studia Sci. Math. Hungar.* **5** (1970), 461–470. *MR* **44** #2664

[16] Toft, B., Two theorems on critical 4-chromatic graphs, *Studia Sci. Math. Hungar.* **7** (1972), 83–89. *MR* **47** #4842

[17] Toft, B., *Graph colouring theory*, Inst. Mat. og. Dat., Odense Universitet, 1987.

[18] Turán, P., Eine Extremalaufgabe aus der Graphentheorie, *Mat. Fiz. Lapok* **48** (1941), 436–452 (in Hungarian with German summary). *MR* **8**–284. See also: On the theory of graphs, *Colloquium Math.* **3** (1954), 19–30. *MR* **15**–976

[19] TUTTE, W. T., Solution to advanced problem No. 4525, *Amer. Math. Monthly* **61** (1954), 532.

[20] ZEIDL, B., Über 4- und 5-chrome Graphen, *Monatsh. Math.* **62** (1958), 212–218. *MR* **20** #6106

(Received March 24, 1990)

DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE
UNIVERSITY OF ALBERTA
632 CENTRAL ACADEMIC BUILDING
EDMONTON, ALBERTA
T6G 2G1
CANADA

Present address of B. Zhou:

DEPARTMENT OF MATHEMATICS
TRENT UNIVERSITY
PETERBOROUGH, ONTARIO
K9J 7B8
CANADA

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 359–362

NOTE ON ADDITIVE FUNCTIONS SATISFYING SOME CONGRUENCE PROPERTY. I

PHAM VAN CHUNG

Let $\mathcal{M}$, $\mathcal{M}^*$, $\mathcal{A}$ and $\mathcal{A}^*$ denote the set of integer-valued multiplicative, completely multiplicative, additive and completely additive arithmetical functions, respectively. We shall denote by $\mathbf{Z}$ resp. $\mathbf{N}$ the set of integers and positive integers.

In 1966 M. V. Subbarao [6] proved that if $f \in \mathcal{M}$ and f satisfies the relation

$$f(n+m) \equiv f(m) (\mathrm{mod}\, n) \tag{1}$$

for every positive integer n and m, then

$$f(n) = n^{\alpha} \qquad (n \in \mathbf{N}) \tag{2}$$

where α is a non-negative integer. In [1] A. Iványi extended this result proving that if $f \in \mathcal{M}$ and (1) holds for a fixed m and for every positive n, then f also has the form (2). For some generalizations of these results we refer to [2], [4] and [5]. Recently, K. Kovács [3] has proved analogous theorems for integer-valued additive functions.

Our purpose in this paper is to give a characterization of those functions $f \in \mathcal{A}$ which satisfy $f(n+M) \equiv C (\mathrm{mod}\, n)$ for fixed $M \in \mathbf{Z}$, $C \in \mathbf{Z}$ and for all $n > \max\{0, -M\}$. We prove the following

THEOREM. *Let M, C be fixed integers and let $f \in \mathcal{A}$. If*

$$f(n+M) \equiv C (\mathrm{mod}\, n) \quad \textit{for all} \quad n > \max\{0, -M\}, \tag{3}$$

then $f \equiv 0$.

PROOF. The theorem will be proved in three cases on $M > 0$, $M = 0$ and $M < 0$.

Case I. We prove our theorem for $M > 0$ in three steps.

1. First we show that, if $f \in \mathcal{A}$ and (3) holds for every $n \in \mathbf{N}$, then

$$C = f(M). \tag{4}$$

1991 *Mathematics Subject Classifications.* Primary 11A25.
Key words and phrases. Characterization of additive functions.

This paper is granted to a one-year scholarship (at the Eötvös University, Department of Algebra and Number Theory) for Viet-Nameses educated in Hungary.

This time we use (3) with n given by M^2st where s and t run over the natural numbers. Since $(M, Mst+1)=1$ we have

$$f(Mst+1)\equiv C-f(M)(\mathrm{mod}\, t) \tag{5}$$

for all $s,t\in N$. Using (5) with t given by $m!$, we get

$$f(Msm!+1)\equiv C-f(M)(\mathrm{mod}\, m)$$

for all natural numbers s,m. For fixed $m>1$, let us summarize these congruences for all s running from 1 to m. We have

$$\sum_{s=1}^{m} f(Msm!+1)\equiv m(C-F(M))(\mathrm{mod}\, m). \tag{6}$$

The left side sum can be written in form $f(MAm+1)$ since $(Mim!+1, Mjm!+1)=1$ for all $i\neq j$. Replacing n by M^2Am in (3) we get

$$f(MAm+1)\equiv C-f(M)(\mathrm{mod}\, m).$$

This by (6) gives

$$\sum_{s=1}^{m} f(Msm!+1)\equiv C-f(M)\equiv 0(\mathrm{mod}\, m)$$

for all $m>1$, i.e. $C=f(M)$. So for $s=1$ (5) gives

$$f(Mm+1)\equiv 0\ (\mathrm{mod}\, m)\quad (m\in \mathbf{N}). \tag{7}$$

2. Let p be a prime which is coprime to M. For any prime $q>p$, and fixed $k\in \mathbf{N}$, there are infinitely many primes $x>y>q$ such that

$$py\equiv 1(\mathrm{mod}\, qM) \tag{8}$$

and

$$p^k x\equiv 1(\mathrm{mod}\, qM). \tag{9}$$

The congruences (8) and (9), by (7), give

$$f(p^{k+1}xy)-f(p^k x)-f(py)\equiv 0(\mathrm{mod}\, q)$$

which implies $q|f(p^{k+1})-f(p^k)-f(p)$. Hence we obtain $f(p^\beta)=\beta f(p)$ for all $\beta\in \mathbf{N}$. By (3), we have

$$f(M)\equiv f(p^\alpha M)=\alpha f(p)+f(M)(\mathrm{mod}\,(p^\alpha-1)),$$

i.e. $p^\alpha - 1 | \alpha f(p)$. So $\alpha \to \infty$ gives

$$f(p^\beta) = f(p) = 0 \text{ for all } p \nmid M, \beta \in \mathbf{N}. \tag{10}$$

3. Let $p^\beta || M$ with $\beta > 0$. For any $0 \leqq \alpha \leqq \beta$ there are infinitely many $q \in \mathbf{N}$ such that, using (10),

$$f(M) \equiv f(qp^\alpha + M) = f\left(q + \frac{M}{p^\alpha}\right) + f(p^\alpha) = f(p^\alpha) (\mathrm{mod}\, q).$$

So we get

$$f(M) = f(p^\alpha) = f(p^0) = 0 \quad \text{if } 0 \leqq \alpha \leqq \beta. \tag{11}$$

For any fixed $\gamma > 0$ and for all $(n, M) = 1$, by (10) and (11), we have

$$0 = f(M) \equiv f(p^\gamma M n) = f(p^{\gamma+\beta}) + f\left(\frac{M}{p^\beta}\right) + f(n) = f(p^{\gamma+\beta}) (\mathrm{mod}\, (p^\gamma n - 1)).$$

Hence $n \to \infty$ gives also

$$f(p^\delta) = 0 \qquad \text{for all } \delta = \gamma + \beta > \beta. \tag{12}$$

So (10), (11) and (12) imply $f \equiv 0$.

Case II. Let us consider the case $M = 0$. We get from (3) that

$$f(n) \equiv C (\mathrm{mod}\, n). \tag{13}$$

Let u be a fixed positive integer. Since there exist infinitely many positive m with $(m, u) = 1$, we obtain from (13) that

$$f(u) + C \equiv f(u) + f(m) = f(um) \equiv C (\mathrm{mod}\, m).$$

So we have $f(u) \equiv 0 (\mathrm{mod}\, m)$ for all $m \in \mathbf{N}$, which implies $f(u) = 0$.

Case III. Finally we assume that $M < 0$. In this case we can apply (3) with n given by $|M|^2 ms$ to get

$$f(|M| ms - 1) \equiv C - f(|M|) (\mathrm{mod}\, m). \tag{14}$$

Similarly, as in Case I.1, applying (14) with odd integer m, we also have

$$0 \equiv \sum_{s=1}^{m} f(|M| sm! - 1) \equiv C - f(|M|) (\mathrm{mod}\, m),$$

i.e.

$$f(|M|) = C.$$

So the choice $s = 1$ in (14) gives

$$f(|M|m-1) \equiv 0 (\mathrm{mod}\, m) \quad \text{for all } m \in \mathbf{N}. \tag{15}$$

If M is even, then it follows from (15) that

$$f(|M|m+1) \equiv f(|M|m+1) + f(|M|m-1) = f(|M|^2 m^2 - 1) \equiv 0 (\mathrm{mod}\, m)$$

since $(|M|m-1, |M|m+1) = 1$. The last congruence gives similarly to Case I that $f \equiv 0$.

It remains to consider the case with M odd. Then (15) implies that

$$f(2|M|m+1) \equiv f(2|M|m+1) + f(2|M|m-1) = f(4M^2 m^2 - 1) \equiv 0 (\mathrm{mod}\, m)$$

and so, similarly as in Case I, $f(n) = 0$ for all positive integers n coprime to 2.

Finally we show $f(2^k) = 0$ $(k \in \mathbf{N})$. Let Q be coprime to $2M$. We get

$$f(2^k Q^s |M|) = f(2^k) + f(Q^s) + f(|M|) = f(2^k)$$

because $f(|M|) = f(Q^s) = 0$ holds. Using (3), replacing n by $(2^k Q^s + 1)|M|$, we get

$$f(2^k) = f[(2^k Q^s + 1)|M| - |M|] \equiv f(|M|) = 0 (\mathrm{mod}\, |M|(2^k Q^s + 1)).$$

Thus $s \to \infty$ gives $f(2^k) = 0$.

ACKNOWLEDGEMENT. I wish to express my gratitude to K. Kovács for her remarks.

REFERENCES

[1] IVÁNYI, A., On multiplicative functions with congruence property, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **15** (1972), 133–137. *MR* **48** #8364

[2] JOÓ, I., Note on multiplicative functions satisfying a congruence property. I–II, *Acta Sci. Math. (Szeged)* **55** (1991), 301–307; *Acta Math. Hungar.* **62** (1993), 163–171.

[3] KOVÁCS, K., On additive functions satisfying some congruence properties, *Period. Math. Hungar.* **23** (1991), 227–231.

[4] PHONG, B. M., Multiplicative functions satisfying a congruence property, *Studia Sci. Math. Hungar.* **26** (1991), 123–128.

[5] PHONG, B. M., Multiplicative functions satisfying a congruence property. V, *Acta Math. Hungar.* **62** (1993), 81–87.

[6] SUBBARAO, M. V., Arithmetic functions satisfying a congruence property, *Canad. Math. Bull.* **9** (1966), 143–146. *MR* **33** #3993

(Received March 25, 1990)

KHOA TOAN
ĐAI HOC SU PHAM I
HANOI
VIETNAM

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 363–374

SOME SUFFICIENT CONDITIONS FOR CONVEXITY OF MULTIVARIATE BERNSTEIN–BÉZIER POLYNOMIALS AND BOX SPLINE SURFACES

MING-JUN LAI

Abstract

A sufficient condition on the B-net of a multivariate polynomial in Bernstein–Bézier representation to guarantee that its graph is a convex surface is presented in this paper. Some further studies on the convexity of trivariate polynomials are also included. As an application, some sufficient conditions for convexity of bivariate box spline surfaces are given. These conditions improve some of the results in [8].

1. Introduction

Let $\mathbf{v}^0, \ldots, \mathbf{v}^s \in \mathbf{R}^s$ be $s+1$ distinct points such that the convex hull $\langle \mathbf{v}^0, \ldots, \mathbf{v}^s \rangle = \left\{ \sum_{i=0}^{s} \lambda_i \mathbf{v}^i : \sum_{i=0}^{s} \lambda_i = 1, \text{ and } \lambda_i \geqq 0 \right\}$ is an s-simplex in $\mathbf{R}^s$, $s \geqq 1$. Let $\lambda_i, i = 0, \ldots, s$, satisfy

$$\mathbf{x} = \sum_{i=0}^{s} \lambda_i \mathbf{v}^i, \qquad \sum_{i=0}^{s} \lambda_i = 1$$

and let $T = \langle \mathbf{v}^0, \ldots, \mathbf{v}^s \rangle$. Then $\lambda = (\lambda_0, \ldots, \lambda_s)$ are the barycentric coordinates of $\mathbf{x}$ with respect to T. It is known that each λ_i is a linear function of $\mathbf{x}$. Thus,

$$p_n(\mathbf{x}) = \sum_{|\alpha|=n} c_\alpha B_\alpha(\lambda)$$

with $B_\alpha(\lambda) = \frac{n!}{\alpha!} \lambda_0^{\alpha_0} \ldots \lambda_s^{\alpha_s}$, $\alpha \in \mathbf{Z}^{s+1}$, is a polynomial of total degree $\leqq n$. It is well-known that any polynomial of total degree $\leqq n$ can be expressed in the above way. Such representation for a polynomial p_n is called Bernstein–Bézier representation, in short, B-form (cf. [1]). The coefficient c_α of p_n is called B-coefficient with respect to T, $\alpha \in \mathbf{Z}_+^{s+1}$ with $|\alpha| = n$. The coefficients c_α together with $\mathbf{x}_\alpha = \frac{1}{|\alpha|} \sum_{i=0}^{s} \alpha_i \mathbf{v}^i$ with $|\alpha| = n$ constitute the B-net of p_n which

---

1980 *Mathematics Subject Classifications* (1985 Revision). Primary 41A15; Secondary 41A63, 26B25, 65D07.

Key words and phrases. Polynomial surfaces, box spline surfaces, conditions for convexity of polynomial surfaces and box spline surfaces.

contain "visible" information about the geometric feature of the surface p_n. This is one of the reasons why polynomials in B-form have been widely used in Computer Aided Geometric Design.

The relation between properties of B-coefficients $c_\alpha, |\alpha| = n$ and the convexity of the polynomial surface p_n have been explored for many years. The first sufficient condition on $\{c_\alpha : |\alpha| = n\}$ to ensure the convexity of bivariate polynomial p_n was given by Chang and Davis in [3]. This condition has been improved and generalized since then for instance in [4], [8] and [10]. See a recent survey [9] on the study of shape preservation property of multivariate Bernstein–Bézier polynomials for further references. We need some necessary notation to discuss the results mentioned above and to conduct our investigations.

Let $A, B \in \mathbf{R}^s$ be two distinct points. Define the derivative D_{A-B} along direction AB by

$$D_{A-B}f(\mathbf{x}) = \lim_{t\to 0} \frac{f(\mathbf{x}+t(A-B)) - f(\mathbf{x})}{t}.$$

We denote $D_{ij} := D_{\mathbf{v}^i - \mathbf{v}^j}, i \neq j, i, j = 0, \ldots, s$. Then it is known that

$$D_{ij}p_n(\mathbf{x}) = n \sum_{|\alpha|=n-1} (c_{\alpha+e^i} - c_{\alpha+e^j}) B_\alpha(\lambda)$$

(cf., e.g., [1]). Furthermore, define the difference operator $\triangle_{ij}$ along the direction $\mathbf{v}^i\mathbf{v}^j$ by

$$\triangle_{ij}c_\alpha = c_{\alpha+e^i} - c_{\alpha+e^j}, \quad i \neq j, \quad i, j = 0, \ldots, s.$$

Chang and Davis' sufficient condition for the convexity of a bivariate polynomial $p_n(\mathbf{x})$ on $\langle \mathbf{v}^0, \mathbf{v}^1, \mathbf{v}^2 \rangle$ is that

$$\begin{aligned} &\triangle_{10}\triangle_{20}c_\alpha \geqq 0, \\ &\triangle_{01}\triangle_{21}c_\alpha \geqq 0, \\ &\triangle_{02}\triangle_{12}c_\alpha \geqq 0, \end{aligned} \tag{1}$$

for all $|\alpha| = n-2$. Later, the following improved conditions were established in [4]

$$\triangle_{10}^2 c_\alpha \geqq 0, \qquad \forall |\alpha| = n-2 \tag{2a}$$

and

$$(\triangle_{10}\triangle_{20}c_\alpha)(\triangle_{01}\triangle_{21}c_\alpha) + (\triangle_{10}\triangle_{20}c_\alpha)(\triangle_{02}\triangle_{12}c_\alpha)$$

$$(\triangle_{01}\triangle_{21}c_\alpha)(\triangle_{02}\triangle_{12}c_\alpha) \geqq 0, \qquad \forall |\alpha| = n-2. \tag{2b}$$

In [8], Dahmen and Micchelli established conditions similar to (1) for the general multivariate case. However, while the precise analog to (1) is shown

in [8] to work still in the trivariate case, an example is given there which demonstrates that this form of conditions does not guarantee convexity in higher dimensional cases.

Let us reapproach this problem to find certain sufficient conditions on the B-coefficients of $p_n(\mathbf{x}) = \sum_{|\alpha|=n} c_\alpha B_\alpha(\lambda)$ to guarantee that p_n is a convex surface on T.

For any direction d in $\mathbf{R}^s$, we can write it as

$$d = \sum_{i=1}^{s} \eta_i(\mathbf{v}^i - \mathbf{v}^0),$$

for some $\eta_i, i = 1, \ldots, s$. Then

$$D_d^2 p_n(\mathbf{x}) = n(n-1) \sum_{|\alpha|=n-2} \sum_{i,j=1}^{s} \eta_i \eta_j \triangle_{i0} \triangle_{j0} c_\alpha B_\alpha(\lambda)$$

$$= n(n-1) \sum_{|\alpha|=n-2} \eta^t C_s(\alpha) \eta B_\alpha(\lambda)$$

where $\eta = (\eta_1, \ldots, \eta_s)^t$ and

$$C_s(\alpha) = \begin{bmatrix} \triangle_{10}^2 c_\alpha & \triangle_{10}\triangle_{20} c_\alpha & \cdots & \triangle_{10}\triangle_{s0} c_\alpha \\ \triangle_{20}\triangle_{10} c_\alpha & \triangle_{20}^2 c_\alpha & \cdots & \triangle_{20}\triangle_{s0} c_\alpha \\ \cdots & \cdots & \cdots & \cdots \\ \triangle_{s0}\triangle_{10} c_\alpha & \triangle_{20}\triangle_{s0} c_\alpha & \cdots & \triangle_{s0}^2 c_\alpha \end{bmatrix}. \tag{3}$$

It is well-known that p_n is convex on T if and only if $D_d^2 p_n(\mathbf{x}) \geqq 0$, $\forall \mathbf{x} \in T$ for any direction d.

A trivial sufficient condition to ensure that p_n is convex on T is that $C_s(\alpha)$ is positive semi-definite for $|\alpha| = n-2$. This condition is also necessary when $n = 2$ and $n = 3$ for any $s \geqq 1$. When $s = 2$, $C_2(\alpha)$ is positive semi-definite if and only if (2) holds.

It is known from linear algebra that $C_s(\alpha)$ is positive semi-definite if and only if $\det(C_k(\alpha)) \geqq 0$, for $1 \leqq k \leqq s$ which is equivalent to saying that all the eigenvalues of $C_s(\alpha)$ are nonnegative. In the present text it is important to find possibly simple conditions for positive semi-definiteness of a matrix which is given explicitly in term of the control coefficients c_α and which are easy to verify. Certainly, condition (1) is simple. But it cannot be simply generalized to the multivariate setting when $s \geqq 4$ as shown in [8].

There is another kind of sufficient conditions proved in [3] that if the piecewise linear interpolant to the B-coefficients of bivariate p_n on T is convex, then p_n is convex on T. This condition was later generalized to any multivariate setting by Dahmen and Micchelli and extended to the box spline surfaces. See [8].

We will prove in the next section a simple sufficient condition on the B-coefficients of a multivariate polynomial p_n with respect to T to ensure that it is convex on T. When $s=2$, our condition is better than (1). When $s=3$, there is no comparison between our sufficient condition with the condition in [8] (see Condition (5) below) which is a generalization of Condition (1). Our condition can also guarantee the convexity of a polynomial surface even if a piecewise linear interpolant to the B-coefficients of a polynomial p_n on T is not convex.

We further study sufficient conditions for the trivariate setting in Section 3 and obtain a sufficient condition which is better than the one in [8]. As an application of our condition in Section 2, we also study the convexity of bivariate box spline surfaces and deduce some sufficient conditions to ensure the convexity of spline surfaces.

2. Multivariate case

Let us begin with a definition of weak* diagonal dominance of a matrix.

DEFINITION. A square matrix $A=(a_{ij})_{n\times n}$ is called weak* diagonally dominant if it satisfies

$$|a_{ii}| \geqq \sum_{j\neq i} |a_{ij}|, \quad i=1,\ldots,s.$$

With this definition we can prove the following

LEMMA 1. *Let $A=(a_{ij})_{n\times n}$ be a symmetric real matrix with nonnegative diagonal elements, i.e., $a_{ii}\geqq 0$, $i=1,\ldots,n$. If A has weak* diagonal dominance, then A is positive semi-definite.*

The proof of this lemma is folklore. We omit here.

Thus, we obtain immediately the following

THEOREM 1. *If the B-coefficients c_α of a polynomial $p_n(\mathbf{x})=\sum_{|\alpha|=n} c_\alpha B_\alpha(\lambda)$ defined on T satisfy*

$$\triangle_{i0}^2 c_\alpha \geqq \sum_{\substack{j=1\\ j\neq i}}^{s} |\triangle_{j0}\triangle_{i0} c_\alpha|, \qquad \forall |\alpha|=n-2,\ i=1,2,\ldots,s, \tag{4}$$

then p_n is convex on T.

PROOF. In view of Lemma 1, Condition (4) implies that $C_s(\alpha)$, as defined in (3), is positive semi-definite. Hence $D_d^2 p_n(\mathbf{x})\geqq 0, \forall \mathbf{x}\in T$ and for any direction d. Thus, p_n is convex on T. This completes the proof of Theorem 1.

Similarly, we can prove the following

COROLLARY. *For a polynomial* $p_n(\mathbf{x}) = \sum_{|\alpha|=n} c_\alpha B_\alpha(\lambda)$ *defined on* T, *if its B-coefficients* c_α*'s satisfy*

$$\Delta_{il}^2 c_\alpha \geqq \sum_{\substack{j=0 \\ j\neq i, j\neq l}}^{s} |\Delta_{il}\Delta_{jl} c_\alpha|, \qquad \forall |\alpha| = n-2,\ i \neq l$$

for some $0 \leqq l \leqq s$, *then* p_n *is convex on* T.

REMARK. It is known that when A is positive semi-definite, $a_{ii} = 0$ implies that $a_{ij} = 0$, and $a_{ji} = 0, \forall j \neq i$. This indicates that weak* diagonal dominance is close to a necessary condition for the positive semi-definiteness of A.

We now consider the special case $s = 2$.

PROPOSITION 1. *Let* p_n *be a bivariate polynomial of total degree* n *defined on* $\langle \mathbf{v}^0, \mathbf{v}^1, \mathbf{v}^2 \rangle$. *If they satisfy Condition* (1), *its B-coefficients satisfy Condition* (4).

PROOF. $\Delta_{10}^2 c_\alpha - \Delta_{10}\Delta_{20} c_\alpha = \Delta_{10}(\Delta_{10} - \Delta_{20}) c_\alpha = \Delta_{10}\Delta_{12} c_\alpha = \Delta_{01}\Delta_{21} c_\alpha \geqq 0$ by Condition (1). Thus, we have $\Delta_{10}^2 c_\alpha \geqq \Delta_{10}\Delta_{20} c_\alpha = |\Delta_{10}\Delta_{20} c_\alpha|$. Similarly, we can show that $\Delta_{20}^2 c_\alpha \geqq |\Delta_{10}\Delta_{20} c_\alpha|$ for $|\alpha| = n - 2$. Thus, $C_2(\alpha)$ has weak*diagonal dominance and hence, c_α satisfies our Condition (4). This completes the proof.

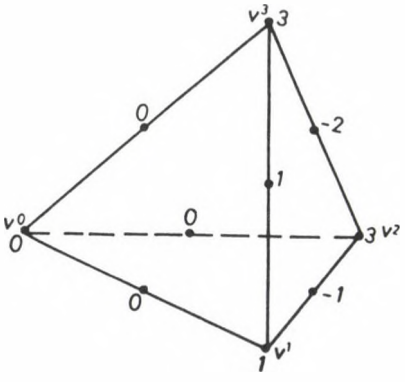


Fig. 1

Let us consider the following quadratic polynomial $p_2(\mathbf{x})$ defined on $\langle \mathbf{v}^0, \mathbf{v}^1, \mathbf{v}^2, \mathbf{v}^3 \rangle$ with zero B-coefficients except for $c_{(0,2,0,0)} = 2$, $c_{(0,0,2,0)} = 3$, $c_{(0,0,0,2)} = 3$, $c_{(0,1,1,0)} = -1$, $c_{(0,0,1,1)} = -2$, $c_{(0,1,0,1)} = 1$ as shown in Figure 1.

It is easy to find that $\triangle_{10}^2 c_{(0,0,0,0)} = 2$, $\triangle_{20}^2 c_{(0,0,0,0)} = \triangle_{30}^2 c_{(0,0,0,0)} = 3$, but $\triangle_{10}\triangle_{20} c_{(0,0,0,0)} = -1$, $\triangle_{10}\triangle_{30} c_{(0,0,0,0)} = 1$ and $\triangle_{20}\triangle_{30} c_{(0,0,0,0)} = -2$. By our Condition (4), the polynomial p_2 is convex on $\langle \mathbf{v}^0, \mathbf{v}^1, \mathbf{v}^2, \mathbf{v}^3 \rangle$. But the condition in [8] which is a generalization of Condition (1) fails to see the convexity of p_2.

However, the following example $q_2(\mathbf{x})$ with zero B-coefficients except $c_{(2,0,0,0)} = 2$, $c_{(0,2,0,0)} = 1$, $c_{(0,0,2,0)} = 1$, and $c_{(0,0,0,2)} = 1$ as shown in the following Figure 2. By the condition in [8], it is convex. But our Condition (4) fails to see that.

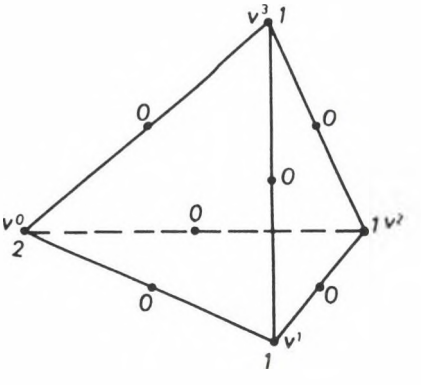


Fig. 2

3. Trivariate case

We now further study the trivariate case. Let

$$P_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad P_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad P_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P_4 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad P_5 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad P_6 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

be positive semi-definite matrices. Then it can be easily shown that

$$C_3(\alpha) = \begin{bmatrix} \triangle_{10}^2 c_\alpha & \triangle_{10}\triangle_{20} c_\alpha & \triangle_{10}\triangle_{30} c_\alpha \\ \triangle_{10}\triangle_{20} c_\alpha & \triangle_{20}^2 c_\alpha & \triangle_{20}\triangle_{30} c_\alpha \\ \triangle_{10}\triangle_{30} c_\alpha & \triangle_{20}\triangle_{30} c_\alpha & \triangle_{30}^2 c_\alpha \end{bmatrix} = \sum_{j=1}^{6} t_j(\alpha) P_j$$

with $t_1(\alpha) = \triangle_{21}\triangle_{31} c_\alpha$, $t_2(\alpha) = \triangle_{02}\triangle_{12} c_\alpha$, $t_3(\alpha) = \triangle_{03}\triangle_{13} c_\alpha$, $t_4(\alpha) = \triangle_{20}\triangle_{30} c_\alpha$, $t_5(\alpha) = \triangle_{02}\triangle_{31} c_\alpha$, and $t_6(\alpha) = \triangle_{12}\triangle_{30} c_\alpha$. Thus we have the following proposition.

PROPOSITION 2. *If* $t_i(\alpha) \geqq 0, \forall i = 1, \ldots, 6$ *and* $|\alpha| = n-2$, *then* $C_3(\alpha)$ *is positive semi-definite for each* $|\alpha| = n-2$ *and hence,* $p_n(\mathbf{x})$ *is convex on* $\langle \mathbf{v}^0, \mathbf{v}^1, \mathbf{v}^2, \mathbf{v}^3 \rangle$.

Furthermore, let

$$P_7 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

be another positive semi-definite matrix. It is easy to see that $P_1 + P_2 + P_3 + P_4 - P_5 - P_6 - P_7 = 0$. Now we have

LEMMA 2. *If* $\min\limits_{1 \leqq i \leqq 4} t_i(\alpha) \geqq \max\limits_{5 \leqq j \leqq 6} \{-t_j(\alpha)\}$, *then* $C_3(\alpha)$ *is positive semi-definite.*

PROOF. We consider two cases.

Case 1. $\min\limits_{1 \leqq i \leqq 4} t_i(\alpha) < 0$.

Let $y_i(\alpha) = y_7(\alpha) + t_i(\alpha), i = 1, 2, 3, 4$ and $y_j(\alpha) = -y_7(\alpha) + t_j(\alpha), 5 \leqq j \leqq 6$ for such $y_7(\alpha) \geqq 0$ that

$$\min_{1 \leqq i \leqq 4} t_i(\alpha) \geqq -y_7(\alpha) \geqq \max_{5 \leqq j \leqq 6} \{-t_j(\alpha)\}.$$

Then $y_i(\alpha) \geqq 0, 1 \leqq i \leqq 6$. Since

$$C_3(\alpha) = \sum_{i=1}^{7} y_i(\alpha) P_i,$$

$C_3(\alpha)$ is positive semi-definite.

Case 2. $\min\limits_{1 \leqq i \leqq 4} t_i(\alpha) \geqq 0$.

Let $\bar{y}_i(\alpha) = -\bar{y}_7(\alpha) + t_i(\alpha), i = 1, 2, 3, 4$ and $\bar{y}_j(\alpha) = \bar{y}_7(\alpha) + t_j(\alpha), j = 5, 6$ for some $\bar{y}_7(\alpha) \geqq 0$ such that

$$\min_{1 \leqq i \leqq 4} t_i(\alpha) \geqq \bar{y}_7(\alpha) \geqq \max_{5 \leqq j \leqq 6} \{-t_j(\alpha)\}.$$

Then $\bar{y}_i(\alpha) \geqq 0, i = 1, \ldots, 7$. Since

$$C_3(\alpha) = \sum_{i=1}^{7} \bar{y}_i(\alpha) P_i,$$

we conclude that $C_3(\alpha)$ is positive semi-definite. Hence, the proof is complete.

COROLLARY (Dahmen and Micchelli). *If the B-coefficients c_α's of p_n satisfy*

$$\triangle_{ik}\triangle_{jk} c_\alpha \geqq 0, \qquad i, j, k = 0, 1, 2, 3, \qquad \forall |\alpha| = n - 2, \tag{5}$$

then $C_3(\alpha)$ is positive semi-definite for all $|\alpha| = n-2$ and hence, p_n is convex.

PROOF. We only need to show that under the assumption (5), $\min\limits_{1 \leqq i \leqq 4} t_i(\alpha) \geqq \max\limits_{5 \leqq j \leqq 6} \{-t_j(\alpha)\}$, that is, $-\triangle_{02}\triangle_{31} c_\alpha \leqq t_i(\alpha), i = 1, 2, 3, 4$ and $-\triangle_{12}\triangle_{30} c_\alpha \leqq t_i(\alpha), i = 1, 2, 3, 4$.

For instance, $-\triangle_{02}\triangle_{31} c_\alpha = \triangle_{02}\triangle_{13} c_\alpha = \triangle_{02}\triangle_{03} c_\alpha + \triangle_{02}\triangle_{10} c_\alpha = = \triangle_{02}\triangle_{03} c_\alpha - \triangle_{02}\triangle_{01} c_\alpha \leqq \triangle_{02}\triangle_{03} c_\alpha = t_4(\alpha)$.

Also, $-\triangle_{02}\triangle_{31} c_\alpha = \triangle_{02}\triangle_{13} c_\alpha = \triangle_{03}\triangle_{13} c_\alpha + \triangle_{32}\triangle_{13} c_\alpha \leqq \triangle_{03}\triangle_{13} c_\alpha = = t_3(\alpha)$ since $\triangle_{32}\triangle_{13} = -\triangle_{23}\triangle_{13} \leqq 0$.

Similarly, we can show that $-\triangle_{20}\triangle_{31} \leqq t_1(\alpha)$ and $\leqq t_2(\alpha)$.

By the same argument we can show that $-\triangle_{12}\triangle_{30} c_\alpha \leqq t_i(\alpha), i = 1, 2, 3, 4$. This completes the proof.

By using Lemma 1 and 2, we immediately get

THEOREM 2. *Let $p_n(\mathbf{x}) = \sum\limits_{|\alpha|=n} c_\alpha B_\alpha(\lambda)$ defined on $\langle \mathbf{v}^0, \mathbf{v}^1, \mathbf{v}^2, \mathbf{v}^3 \rangle$. Suppose that for each α with $|\alpha| = n - 2$, $C_3(\alpha)$ satisfies either $\triangle_{i0}^2 c_\alpha \geqq \geqq \sum\limits_{\substack{j \neq i \\ j=1}}^{3} |\triangle_{i0}\triangle_{j0} c_\alpha|, i = 1, 2, 3$ or $\min\limits_{1 \leqq i \leqq 4} t_i(\alpha) \geqq \max\limits_{5 \leqq j \leqq 6} t_j(\alpha)$. Then $p_n(\mathbf{x})$ is convex on $\langle \mathbf{v}^0, \mathbf{v}^1, \mathbf{v}^2, \mathbf{v}^3 \rangle$.*

We conclude this section with the following example. A quadratic polynomial $r_2(\mathbf{x})$ with given B-coefficients: $c_{(2,0,0,0)} = c_{(1,1,0,0)} = c_{(0,2,0,0)} = = c_{(0,0,1,1)} = 0$, $c_{(1,0,0,1)} = c_{(1,0,1,0)} = c_{(0,1,0,1)} = c_{(0,1,1,0)} = 1$, and $c_{(0,0,2,0)} = = c_{(0,0,0,2)} = 2$. See Figure 3 for reference. Then it is easy to find out that $t_1(0) = -2$, $t_2(0) = 0$, $t_3(0) = 0$, $t_4(0) = -2$ and $t_5(0) = 2$, $t_6(0) = 2$. By using Lemma 2, $C_3(0)$ is positive semi-definite. But, $\triangle_{20}\triangle_{30} c_{(0,0,0,0)} = -2$ which violates (5). Thus, (5) fails to see the convexity of $r_2(\mathbf{x})$.

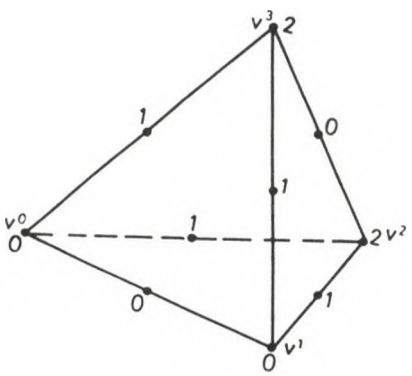


Fig. 3

4. Bivariate box spline case

In this section, we shall apply the sufficient condition discussed in Section 2 to the study on the convexity of bivariate box spline surfaces.

First, let us recall some necessary definition and notations. Let $X_n = \{\mathbf{x}^1, \ldots, \mathbf{x}^n\} \subset \mathbf{Z}^2 \backslash \{0\}$ be a direction set. Then the box spline $B(\cdot; X_n)$ is defined by requiring that

$$\int_{\mathbf{R}^2} f(\mathbf{x}) B(\mathbf{x}; X_n) d\mathbf{x} = \int_{[-1/2,1/2]^n} f\left(\sum_{i=1}^{n} \lambda_i \mathbf{x}^i\right) d\lambda$$

holds for all continuous functions f on $\mathbf{R}^2$. Fix a box spline $B(\cdot; X_n)$. The box spline surface $S(\cdot; X_n)$ is defined by

$$S(\cdot; X_n) = \sum_{\alpha \in \mathbf{Z}^2} f(\alpha) B(\cdot - \alpha; X_n).$$

For properties of box spline and box spline surfaces, see, for instance, [2] and [7].

Our purpose in this section is to find some sufficient conditions on $f(\alpha)$ to ensure the convexity of $S(\cdot; X_n)$. It is known that the convexity of $f(\cdot)$ does not guarantee that $S(\cdot; X_n)$ is convex (cf. [8]). It was also shown in [8] that if the piecewise linear function s_0 on a three-directional mesh

interpolating $f(\alpha), \alpha \in \mathbf{Z}^2$ is convex, then $S(\cdot; X_n)$ is convex provided X_n contains e^1, e^2, and $e^1 + e^2$, where $e^1 = (1,0)$ and $e^2 = (0,1)$. (The results in [8] are formulated for the general s-variate case.) Their results is based on the following

PROPOSITION 3 (Dahmen and Micchelli). *If $S(\cdot; V_m)$ is convex, then $S(\cdot; V_m \cup \{\mathbf{v}\})$ is convex for any direction* $\mathbf{v}$.

However, a fact is that even if the piecewise linear interpolant s_0 is not convex, $S(\cdot; X_n)$ may also be convex. Thus, we need further study to find finer conditions to ensure the convexity of $S(\cdot; X_n)$.

Let $Y_3 = \{e^1, e^2, e^1 + e^2\}, Y_4 = \{e^1, e^2, e^1 + e^2, e^1 - e^2\}$ and $Y_5 = \{e^1, e^1, e^2, e^2, e^1 + e^2\}$.

We first consider $S(\cdot; X_n)$ with $Y_5 \subset X_n$. The B-coefficients of $B(\cdot; Y_5)$ can be found in [5] or [6]. Let U_i and L_i be the upper and lower triangles of $[i, i+1] \times [j, j+1]$, respectively. After the computation of all second differences along directions e^2 and $e^1 + e^2$ of the B-coefficients of $S(\cdot; Y_5)|_{U_i}$ and all second differences along directions e^1 and $e^1 + e^2$ of the B-coefficients of $S(\cdot; Y_5)|_{L_i}$, we apply Theorem 1 and Proposition 3 to get

THEOREM 3. *For any X_n which contains Y_5, $S(\cdot; X_n)$ is convex if $f_{ij} := f(i,j), (i,j) \in \mathbf{Z}^2$ satisfy the following conditions*

$$\min\{f_{i,j+1} - 2f_{ij} + f_{i,j-1}, f_{i+1,j} - f_{i-1,j} + f_{i,j+1} - f_{i,j-1} - 2(f_{ij} - f_{i-1,j-1})\} \geqq |f_{i,j+1} - f_{ij} - f_{i-1,j} + f_{i-1,j-1}|,$$

$$\min\{f_{i,j+1} - 2f_{ij} + f_{i,j-1}, 2(f_{i+1,j+1} - f_{ij}) - (f_{i+1,j} - f_{i-1,j}) - (f_{i,j+1} - f_{i,j-1})\} \geqq |f_{i+1,j+1} - f_{i+1,j} - f_{ij} + f_{i,j-1}|,$$

$$\min\{f_{i-1,j+1} - 2f_{i-1,j} + f_{i-1,j-1}, f_{i+1,j+1} - 2f_{ij} + f_{i-1,j-1}\} \geqq |f_{i,j+1} - f_{ij} - f_{i-1,j+1} + f_{i-1,j}|,$$

and

$$\min\{f_{i+1,j} - 2f_{ij} + f_{i-1,j}, f_{i+1,j} - f_{i-1,j} + f_{i,j+1} - f_{i,j-1} - 2(f_{ij} - f_{i-1,j-1})\} \geqq |f_{i+1,j} - f_{ij} - f_{i,j-1} + f_{i-1,j-1}|,$$

$$\min\{f_{i+1,j-1} - 2f_{ij-1} + f_{i-1,j-1}, f_{i+1,j+1} - 2f_{ij} + f_{i-1,j-1}\} \geqq |f_{i+1,j} - f_{ij} - f_{i,j-1} + f_{i-1,j-1}|,$$

$$\min\{f_{i+1,j} - 2f_{ij} + f_{i-1,j}, 2(f_{i+1,j+1} - f_{ij}) - (f_{i+1,j} - f_{i-1,j}) - (f_{i,j+1} - f_{i,j-1})\} \geqq |f_{i+1,j+1} - f_{ij+1} - f_{ij} + f_{i-1,j}|$$

for $(i,j) \in \mathbf{Z}^2$.

REMARK. When $S(\cdot; Y_3)$ is convex, that is, $f_{ij}, (i,j) \in \mathbf{Z}^2$ satisfy that

$$f_{i+1,j} - 2f_{ij} + f_{i-1,j} \geqq 0,$$

$$f_{i,j+1} - 2f_{ij} + f_{i,j-1} \geqq 0,$$

and

$$f_{i+1,j+1} - f_{ij} - f_{i,j+1} + f_{i-1,j} \geqq 0,$$
$$f_{i,j+1} - f_{ij} - f_{i+1,j+1} + f_{i+1,j} \geqq 0,$$
$$f_{i+1,j+1} - f_{ij} - f_{i+1,j} + f_{i,j-1} \geqq 0.$$

It is easy to verify that $f_{ij}, (i,j) \in \mathbf{Z}^2$ satisfy the condition of Theorem 3, too. Hence, $S(\cdot; X_n)$ is convex provided $Y_5 \subset X_n$.

Next, we consider $S(\cdot; X_n)$ with X_n containing Y_4. The B-coefficients of $B(\cdot; Y_4)$ can be found in [5] or [6]. After the computation of all the second differences of the B-coefficients of $S(\cdot; Y_4)$ restricted to each of the four triangles of $[i, i+1] \times [j, j+1]$, we apply Theorem 1 and Proposition 3 to obtain the following

THEOREM 4. *For any X_n containing Y_4, $S(\cdot; X_n)$ is convex if $f_{ij} := f(i,j), (i,j) \in \mathbf{Z}^2$ satisfy the following condition*

$$(f_{i+1,j+1} - 2f_{i,j+1} + f_{i-1,j+1}) + (f_{i+1,j} - 2f_{ij} + f_{i-1,j}) \geqq$$
$$\geqq |f_{i+1,j+1} - f_{i,j+1} - f_{ij} + f_{i-1,j}|,$$
$$(f_{i+1,j+1} - 2f_{i+1,j} + f_{i+1,j-1}) + (f_{i,j+1} - 2f_{ij} + f_{i,j-1}) \geqq$$
$$\geqq |f_{i+1,j+1} - f_{i+1,j} - f_{ij} + f_{i,j-1}|,$$

and

$$2(f_{i+1,j+1} - f_{ij}) - (f_{i+1,j} - f_{i-1,j}) - (f_{i,j+1} - f_{i,j-1}) \geqq$$
$$2\max\{|f_{i+1,j+1} - f_{i,j+1} - f_{ij} + f_{i-1,j}|, |f_{i+1,j+1} - f_{i+1,j} - f_{ij} + f_{i,j-1}|\},$$
$$(f_{i,j+1} - f_{ij-1}) + (f_{i+1,j} - f_{i-1,j}) - 2(f_{i,j} - f_{i-1,j-1}) \geqq$$
$$2\max\{|f_{i+1,j} - f_{i,j} - f_{ij-1} + f_{i-1,j-1}|, |f_{i,j+1} - f_{i-1,j} - f_{ij} + f_{i-1,j-1}|\}$$

for all $(i,j) \in \mathbf{Z}^2$.

REFERENCES

[1] BOOR, C. DE, *B*-form basics, *Geometric modeling*, G. Farin ed., SIAM Publication, Philadelphia, PA, 1987, 131–148. *MR* **89b**:65010

[2] BOOR, D. DE and HÖLLIG, K., B-splines from parallelepipeds, *J. Analyse Math.* **42** (1982/83), 99–115. *MR* **86d**:41008

[3] CHANG, G. and DAVIS, P. J., The convexity of Bernstein polynomials over triangles, *J. Approx. Theory* **40** (1984), 11–28. *MR* **85c**:41001

[4] CHANG, G. and FENG, Y. Y., An improved condition for the convexity of Bernstein–Bézier surfaces over triangles, *Comput. Aided Geom. Design* **1** (1984), 279–283.

[5] CHUI, C. K., *Multivariate splines*, CBMS–NSF Reg. Conf. Ser. Appl. Math. **54**, SIAM Publication, Philadelphia, 1988. *Zbl* **687**.41018

[6] CHUI, C. K. and LAI, M. J., Computation of box splines and B-splines on triangulation of non-uniform rectangular partitions, Proceedings of China–U.S. Joint Conference on Approximation Theory (Hangzhou, 1985), *Approx. Theory Appl.* **3** (1987), 37–62. *MR* **89e**:65012

[7] DAHMEN, W. and MICCHELLI, C. A., Recent progress in multivariate splines, *Approximation Theory, IV*, (College Station, Tex., 1983), Academic Press, New York, 1983, 27–121. *MR* **85h**:41013

[8] DAHMEN, W. and MICCHELLI, C. A., Convexity of multivariate Bernstein polynomials and box spline surfaces, *Studia Sci. Math. Hungar.* **23** (1988), 265–287. *MR* **90g**:41005

[9] GOODMAN, T. N. T., Shape preserving representations, *Mathematical methods in computer aided geometric design* (Oslo, 1988), T. Lyche and L. L. Schumaker eds., Academic Press, 1989, 333–351. *MR* **91a**:65031

[10] ZHENG, W. and LIU, Q., An improved condition for the convexity and positivity of Bernstein–Bézier surfaces over triangles, *Comput. Aided Geometric Design* **5** (1988), 269–275.

(Received April 5, 1990)

DEPARTMENT OF MATHEMATICS
THE UNIVERSITY OF UTAH
SALT LAKE CITY, UT 84112
U.S.A.

Present address:

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF GEORGIA
ATHENS, GA 30602
U.S.A.

e-mail: mjlai@wiener.math.uga.edu

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 375–378

COVERING A PLANE CONVEX BODY WITH NEGATIVE HOMOTHETICAL COPIES

M. LASSAK and É. VÁSÁRHELYI

Covering a plane convex body with positive homothetical copies is considered by many authors (for references see for survey paper [3]). On the other hand, every plane convex body C can be covered by a homothetical copy of C of ratio -2 (see [7]). A natural question emerges about the covering of C with more than one negative copy.

THEOREM. *Every plane convex body C can be covered by two homothetical copies of ratio $-\sqrt{2}$, by three copies of ratio -1, by four copies of a ratio greater than -1.*

PROOF. By Lemma 3 of [5] we can inscribe in C a centrally symmetric hexagon $H = h_1h_2h_3h_4h_5h_6$ with

$$\overrightarrow{h_1h_4} = \left(1 + \frac{1}{2}\sqrt{2}\right)\overrightarrow{h_2h_3}. \tag{1}$$

Let o denote the centre of H. Three of the lines containing the sides of H bound a triangle $V = s_1s_3s_5$ containing H, and three other a triangle $W = s_2s_4s_6$; the notation is chosen such that s_1, h_2, h_3, s_3 are successive points on a line, and that s_{i+3} is symmetric to s_i with respect to o for $i = 1, 2, 3$. Let $S = V \cup W$. Since H is inscribed in C, the convexity of C implies $C \subset S$. Let p_1 denote the intersection of segments h_1s_3 and h_4s_1. Let p_2 be the intersection of segments h_1s_4 and h_4s_6. From (1) it results that $\overrightarrow{s_3s_1} = -\sqrt{2}\overrightarrow{h_1h_4} = \overrightarrow{s_4s_6}$. Consequently, S is covered by the union of homothetical copies of H with centers p_1, p_2 and ratio $-\sqrt{2}$. Thus the inclusions $H \subset C$ and $C \subset S$ show that C is covered by two copies of ratio $-\sqrt{2}$.

It is well known [1] that an affine image $K = k_1k_2k_3k_4k_5k_6$ of a regular hexagon can be inscribed in C. Three lines containing the sides of H bound a triangle X and the other three a triangle Y. Let $Z = X \cup Y$. Since H is inscribed in C, from the convexity of C we obtain $C \subset Z$. Observe that Z is covered by the union of the homothetical copies of H with ratio -1 and centers in the midpoints of the sides of the triangle $k_1k_3k_5$. This and the

1991 *Mathematics Subject Classifications.* Primary 52A45.
Key words and phrases. Homothetical covering, convex bodies.

Akadémiai Kiadó, Budapest

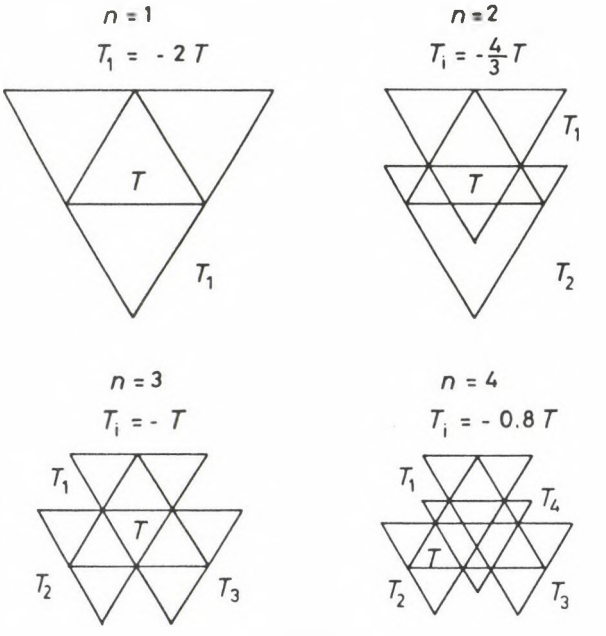


Fig. 1

inclusions $H \subset C$ and $C \subset Z$ imply that C is covered by three homothetical copies of C of ratio -1.

The last statement of Theorem is obvious when C is a triangle (see Fig. 1). Consider the case when C is not a triangle. Then there exists a parallelogram P and its image Q under a homothety with positive ratio $\omega < 2$ such that the inclusions $P \subset C \subset Q$ hold true [6]. Of course, four homothetical copies of P of the ratio $-\frac{1}{2}\omega$ cover Q. From the inclusions $P \subset C$ and $C \subset Q$ we see that C can be covered by four homothetical copies of the ratio $-\frac{1}{2}\omega > -1$.

The proof is complete.

Denote by $g_m(C)$ the greatest negative ratio of m homothetical copies of C whose union covers C. A short consideration analogical to that on p. 163 of [4] shows that the number $g_m(C)$ exists. Our Theorem and the fact quoted before that every plane convex body C can be covered by a homothetical copy of ratio -2 can be formulated as follows:

$$(2) \qquad g_1(C) \geqq -2, \quad g_2(C) \geqq -\sqrt{2}, \quad g_3(C) \geqq -1, \quad g_4(C) > -1.$$

CONJECTURE. *For every plane convex body C we have $g_2(C) \geqq -\frac{4}{3}$ and $g_4(C) \geqq -\frac{4}{5}$.*

Proposition formulated below shows that the estimates for $g_1(C)$ and $g_3(C)$ in the Theorem, and for $g_2(C)$ and $g_4(C)$ in the Conjecture cannot be improved.

PROPOSITION. *For every triangle T we have* $g_1(T) = -2$, $g_2(T) = -\frac{4}{3}$, $g_3(T) = -1$ *and* $g_4(T) = -\frac{4}{5}$.

PROOF. It is sufficient to consider the regular triangle of unit side. From Fig. 1 we see that

$$g_1(T) \geqq -2, \quad g_2(T) \geqq -\frac{4}{3}, \quad g_3(T) \geqq -1 \text{ and } g_4(T) \geqq -\frac{4}{5}. \tag{3}$$

If $-2 \leqq \lambda \leqq -\frac{1}{2}$, then by a_λ we denote the maximum of the length of the part of the boundary of T which can be covered by a homothetical copy of T with ratio λ having nonempty intersections with all sides of T. If $-1 \leqq \lambda \leqq 0$, then by b_λ we mean the maximum of the length of the part of the boundary of T which can be covered by a homothetical copy of T with ratio λ disjoint with at least one side of T. We omit a simple consideration which shows that

$$a_\lambda = -1 - 2\lambda \text{ and } b_\lambda = -\lambda. \tag{4}$$

In Parts (i)–(iv) we show that the four inequalities in (3) can be replaced by equalities.

(i) From (4) we have $a_\lambda < 3$ for $\lambda > -2$, and thus $g_1(T) = -2$.

(ii) Consider the covering of T by two homothetical copies Y and Z of T with negative ratio λ of homothety. For instance, let Y contain two vertices of T. The set $T \setminus Y$ is a homothetical copy of T with ratio at least $2 + \lambda$. From $T \setminus Y \subset Z$ and from the equality $g_1(Z) = -2$ shown in (i) we conclude that $\lambda \leqq -2(2 + \lambda)$. Consequently, $\lambda \leqq -\frac{4}{3}$. By the second inequality in (3) we obtain $g_2(T) = -\frac{4}{3}$.

(iii) From $a_\lambda < 1$ and $b_\lambda < 1$ for $\lambda > -1$, and from the third inequality in (3) we obtain $g_3(T) = -1$.

(iv) Let T be covered by a family $\mathcal{T}$ of four homothetical copies T_1, T_2, T_3, T_4 of T with a negative ratio $\lambda \geqq -\frac{4}{5}$ of homothety. Thanks to $g_4(T) \geqq -\frac{4}{5}$ given in (3), it is sufficient to show that $\lambda \leqq -\frac{4}{5}$.

Observe that if at least one triangle from $\mathcal{T}$ has nonempty intersection with every side of T, then $\lambda \leqq -\frac{1}{2}$. In this case from (4) and from $a_\lambda \leqq b_\lambda$ for $-\frac{4}{5} \leqq \lambda \leqq -\frac{1}{2}$ we obtain $3(-\lambda) + (-1 - 2\lambda) \geqq 3$. Thus $\lambda \leqq -\frac{4}{5}$.

Consider the opposite possibility when every triangle from $\mathcal{T}$ is disjoint with a side of T. Then a side S of T is covered by two triangles from $\mathcal{T}$, say, T_1 and T_2. Let the notation be chosen such that the length μ of $S \subset T_1$ is not smaller than the length of $S \cap T_2$. Denote by v the vertex of T which is not an endpoint of S. Consider two cases.

If $\mu > \frac{3}{5}$, then T_1 has empty intersection with the homothetical copy W_1 of T with the ratio $\frac{8}{5} + \lambda$ and the centre v. Since $W_1 \subset T_2 \cup T_3 \cup T_4$, and by the equality $g_3(W_1) = -1$ established in (iii), we conclude that $\lambda \leqq -\frac{8}{5} - \lambda$. Consequently, $\lambda \leqq -\frac{4}{5}$.

Let $\mu \leqq \frac{3}{5}$. By the choice of T_1 we have $1-\mu \leqq \mu$. Denote by W_2 the homothetical copy of T with the ratio $2-\mu+\lambda$ and the centre v. Observe that the interior of W_2 is disjoint with $T_1 \cup T_2$. By (ii) we have $g_2(W_2) = -\frac{4}{3}$ and consequently, $\lambda \leqq -\frac{4}{3}(2-\mu+\lambda)$. This and $\mu \leqq \frac{3}{5}$ imply $\lambda \leqq -\frac{4}{5}$.

The proof of the Proposition is complete.

Considering the areas of a triangle T and of the negative copies covering T, and using the result of [8] we conclude that

$$-2/\sqrt{n} \leqq g_n(T) \leqq -1/\sqrt{n}$$

for every natural n. From [2] we get

$$\lim_{n\to\infty} \sqrt{n} g_n(C) \geqq -\sqrt{3/2}$$

for every plane convex body C, with equality only when C is a triangle.

REFERENCES

[1] BESICOVITCH, A. S., Measure of asymmetry of convex curves, *J. London Math. Soc.* **23** (1948), 237–240. *MR* **10**–320

[2] FÁRY, I., Sur la densité des réseaux de domaines convexes, *Bull. Soc. Math. France* **78** (1950), 152–161. *MR* **12**–526

[3] LASSAK, M., Covering plane convex bodies with smaller homothetical copies, *Intuitive Geometry* (Siófok, 1985), Colloq. Math. Soc. J. Bolyai, Vol. 48, North-Holland, Amsterdam–New York, 1987, 331–337. *MR* **88i**:52023

[4] LASSAK, M., Covering a plane convex body by four homothetical copies with the smallest positive ratio, *Geom. Dedicata* **21** (1986), 151–167. *MR* **88c**:52013

[5] LASSAK, M., Approximation of plane convex bodies by centrally symmetric bodies, *J. London Math. Soc.* (2) **40** (1989), 369–377. *MR* **91a**:52001

[6] LASSAK, M., Approximation of convex bodies by parallelotopes, *Bull. Polish Acad. Sci. Math.* (to appear).

[7] NEUMANN, B. H., On some affine invariants of closed convex regions, *J. London Math. Soc.* **14** (1939), 262–272. *Zbl* **26**,359

[8] VÁSÁRHELYI, É., Über eine Überdeckung mit homothetischen Dreiecken, *Beiträge Algebra Geom.* **17** (1984), 61–70. *Zbl* **554**.52011

(Received April 20, 1990)

INSTYTUT MATEMATYKI I FIZYKI
ATR
PL–85-790 BYDGOSZCZ
POLAND

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
GEOMETRIAI TANSZÉKE
RÁKÓCZI ÚT 5
H–1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 379–386

HERMITE-FEJÉR INTERPOLATIONS OF HIGHER ORDER. IV

R. SAKAI and P. VÉRTESI

This paper is a continuation (or, the second part) of [1]. That means, all the notations, necessary definitions and theorems can be found in the four parts of [1]. Here we refer to them without any further explanation.

5. Further results

5.1. Theorems 3.1–3.3 dealt with even derivatives of $\ell_k^s(x)$ and $e_{2t,k}$. Here we state theorems on $(\ell_k^s)^{(2j+1)}$ and $e_{2t+1,k}$. By previous notations (cf. Part 2) we have

THEOREM 5.1. *For arbitrary real s with $|s| \leqq M$ ($M \geqq 2$, arbitrary fixed integer), we can write, uniformly in n, k and j,*

(5.1)
$$(\ell_k^s)^{(2j+1)} = \left\{ \left(\ell_{kn}^{(\alpha,\beta)}(x) \right)^s \right\}^{(2j+1)}_{x=x_{kn}^{(\alpha,\beta)}} = \frac{(-1)^j}{\sin^2 \vartheta_k} \left(\frac{n}{\sin \vartheta_k} \right)^{2j} \{\pi_{j+1}(s) + q_{2j+1}(s)\}$$

if $j = 0, 1, \dots, \left[\frac{M}{2}\right] - 1$, $\pi_i(s), q_i(s) \in \mathcal{P}_i$, they depend on k and n and satisfy relations

(5.2)
$$\begin{cases} \pi_j(0) = 0, \quad \pi_1(s) = d(1,k)s \\ \pi_j(s+1) = \sum_{i=0}^{j-1} \binom{2j-1}{2i} d(j-i,k) p_i(s) + \frac{1}{2j} \sum_{i=0}^{j-1} \binom{2j}{2i+1} \pi_{i+1}(s), \end{cases}$$

(5.3)
$$\begin{cases} \left| \frac{d^t q_{2j-1}(s)}{ds^t} \right| \leqq |\varepsilon_{kn}|, \qquad t = 0, 1, \dots, j, \\ q_{2j-1}(0) = q_1(s) = 0. \end{cases}$$

Using Theorem 5.1 we express $\pi_j(s)$ in a more compact form. Namely

THEOREM 5.2. *With previous conditions and notations, for $j = 1, 2, \dots, \left[\frac{M}{2}\right]$*

(5.4)
$$\pi_j(s) = (2j-1) \left\{ \frac{\alpha - \beta}{2} s + \left(\frac{\alpha + \beta}{2} s + s + j - 1 \right) x_{kn} \right\} p_{j-1}(s), \quad |s| \leqq M.$$

1991 *Mathematics Subject Classification.* Primary 41A05; Secondary 41A10.
Key words and phrases. Hermite–Fejér interpolation, Jacobi polynomials.

By (5.4) we can get statements saying $\pi_j(s) \neq 0$ for certain j and s. We give two examples.

COROLLARY 5.3. *Let $\alpha \geqq \beta > -1$, $M \geqq 2$ and $0 < \varepsilon < 1$ be arbitrary fixed. Then with proper $c_0 > 0$ and $\delta_1 > 0$, for $t = 1, 2, \ldots, m$*

$$|\pi_j(t)| = |\pi_j(t, x_{kn})| \geqq c_0 \text{ if } \begin{cases} a)\ x_{kn} \in [\varepsilon, 1), & n \geqq n_0, \text{ or} \\ b)\ x_{kn} \in (-1, -\delta_1], & n \geqq n_0. \end{cases} \tag{5.5}$$

Here, as above, $j = 1, 2, \ldots, \left[\frac{M}{2}\right]$. Analogous statement holds when $\alpha \leqq \beta$.

The next complicated looking example is very useful in investigating the process H_{nm}.

COROLLARY 5.4. *Let $\alpha, \beta > -1$ and let $m \geqq 2$ be even. Then*

$$\left|\pi_{\frac{m}{2}}(-m)\right| = \left|\pi_{\frac{m}{2}}(-m, x_{kn})\right| \geqq c_0 > 0, \quad n \geqq n_0, \tag{5.6}$$

if any of the following conditions holds.

(i) $\alpha \geqq B_m$ *and* $0 < a_1 \leqq x_k < 1$,
(ii) $\alpha = A_m - \delta_2$ *and* $0 < a_2 \leqq x_k < 1$,
(iii) $\beta = A_m - \delta_3$ *and* $-1 < x_k \leqq a_3 < 0$,
(iv) $A_m \leqq \alpha = \beta + \frac{2}{m} + \delta_4 < B_m$ *and* $-1 < x_k \leqq a_n < 0$.

Here $A_m = -\frac{1}{2} - \frac{2}{m}$, $B_m = -\frac{1}{2} + \frac{1}{m}$, $\delta_2, \delta_3, \delta_4 > 0$, the numbers c_0, a_1, a_2, a_3, a_4 are properly chosen. Using symmetry, analogous statements hold when $\beta \geqq \alpha$.

DEFINITION. Let

$$Q_j(s) = (-1)^{j+1}\pi_j(-s), \quad j = 1, 2, \ldots, \left[\frac{m}{2}\right], \quad |s| \leqq m. \tag{5.7}$$

Then clearly $(-1)^{j+1}Q_j(-s) = \pi_j(s)$. By Q_j we can formulate our main relation.

5.2. THEOREM 5.5. *For arbitrary fixed $\alpha, \beta > -1$, $m \geqq 2$, we have if $1 \leqq k \leqq n$, $n \geqq 2, \ldots$,*

$$e_{2t+1,knm} = \frac{Q_{t+1}(m) + \varepsilon_{kn}}{(2t+1)! \sin^2 \vartheta_k} \left(\frac{n}{\sin \vartheta_k}\right)^{2t} \qquad 0 \leqq t \leqq \left[\frac{m}{2}\right] - 1. \tag{5.8}$$

5.3. Now we give an interesting application of the previous statement. By Theorem 2.1 in [2] and [3; 2, Remarks 2] we have

Let $-1 < a < 1$, $\alpha, \beta > -1$ and let $m = 2, 4, \ldots$ be fixed even. Then

$$\lim_{n \to \infty} \|H_{nm}^{(\alpha,\beta)}(f, x) - f(x)\|_{[a,1]} = 0 \qquad \forall f \in C \tag{5.9}$$

if

$$\alpha \in [A_m, B_m),\ \ \beta \geqq A_m \ \text{and}\ \alpha - \beta \leqq 2/m. \tag{5.10}$$

The sharpness of condition (5.10) is showed by a slight improvement of [2, Theorem 2.3]:

Let $\alpha, \beta > -1$ and $m \geqq 2$, even. If (5.10) *is violated then* (5.9) *does not hold (in the sense that there exist an interval $[a, 1]$ and a function $f \in C$ with $\overline{\lim}_{n\to\infty} \|H_{nm}^{(\alpha,\beta)}(f,x) - f(x)\|_{[a,1]} > 0$) supposing*

$$|e_{m-1,knm}| \sim \frac{1}{\sin^2 \vartheta_k} \left(\frac{n}{\sin \vartheta_k}\right)^{m-2}, \quad K \geqq c_0,\ \ x_{kn} \in I, \tag{5.11}$$

where the interval I is defined by the violating condition (that means, if $\alpha = A_m - \delta_2$, say, then $I = [a_2, 1)$ (cf. Corollary 5.2).

So the sharpness of (5.10) depends on the validity of (5.11). However, by (5.8), (5.11) holds true whenever $|Q_{\frac{m}{2}}(m, x_{kn})| \geqq c_0 > 0$. Using $|\pi_{\frac{m}{2}}(-m)| = |Q_{\frac{m}{2}}(m)|$ (see (5.7) and Corollary 5.4), we get that (5.10) *is sharp or, with other words, conditions* (5.9) *and* (5.10) *are equivalent.*

Using symmetry we can prove theorems on $\|H_{nm}^{(\alpha,\beta)}(f,x) - f(x)\|_{[-1,b]}$.

5.4. Another similar application corresponds to the mean convergence of $H_{nm}^{(\alpha,\beta)}(f,x)$ (cf. [3]). We omit the details.

6. Proofs

6.1. PROOF OF THEOREM 5.1. First we give the definitions of the polynomials, $\pi_j(s)$ and $q_{2j-1}(s)$ (k and n fixed). Arguing as at 4.1, for sake of simplicity we consider only the special case $j = 1$. By (4.2) and (2.9)

$$(\ell_k^s)''' = (s)_3 \frac{(d(1,k) + \varepsilon_k(1))^3}{\sin^6 \vartheta_k} - (s)_2 \frac{d(1,k) + \varepsilon_k(1)}{\sin^2 \vartheta_k} \frac{n^2}{\sin^2 \vartheta_k}(1 + \varepsilon_k(2)) -$$

$$- \frac{s}{\sin^2 \vartheta_k} \left(\frac{n}{\sin \vartheta_k}\right)^2 (d(2,k) + \varepsilon_k(3)) =$$

$$= -\frac{1}{\sin^2 \vartheta_k} \left(\frac{n}{\sin \vartheta_k}\right)^2 \left\{(s)_2 d(1,k) + s d(2,k) - (s)_3 \frac{(d(1,k) + \varepsilon_k(1))^3}{n^2 \sin^2 \vartheta_k} + \right.$$

$$\left. + (s)_2 (\varepsilon_k(1) + d(1,k)\varepsilon_k(2) + \varepsilon_k(1)\varepsilon_k(2)) + s\varepsilon_k(3)\right\},$$

where the "main part" of $\{\dots\}$ is $\pi_2(s)$, i.e. $\pi_2(s) = (s)_2 d(1,k) + s d(2,k)$ while $q_3(s) := \{\dots\} - \pi_2(s)$. By $(n \sin \vartheta_k)^{-2} \leqq c(\alpha,\beta) K^{-2}$ we get $|q_3^{(t)}(s)| \leqq \leqq |\varepsilon_k|$. Analogous definitions give $\pi_j(s)$ and $q_{2j-1}(s)$, $j \geqq 2$. As for q_3, we

obtain $|q_{2j+1}^{(t)}| \leqq |\varepsilon_k|$. Relations $\pi_1(s) = d(1,k)s$ and $q_1(s) = 0$ come from $(\ell_k^s)' = s\ell_k' = \frac{s}{2}\frac{P_n''(x_k)}{P_n'(x_k)} = sd(1,k)$ (cf. (2.3), (2.8) with $s=2$ and (5.1)), while $\pi_j(0) = q_{2j-1}(0) = 0$ is obvious.

Finally by

$$(\ell_k^{s+1})^{(2j-1)} = (\ell_k^s \ell_k)^{(2j-1)} = \sum_{i=0}^{j-1} \binom{2j-1}{2i} (\ell_k^s)^{(2i)} \ell_k^{(2j-2i-1)} +$$

$$+ \sum_{i=0}^{j-1} \binom{2j-1}{2i+1} (\ell_k^s)^{(2i+1)} \ell_k^{(2j-2i-2)}$$

we get (5.2) considering (2.9), (3.1), (5.1), (2.9) and the relation $\binom{2j-1}{2i+1}\frac{1}{2j-2i-1} = \frac{1}{2j}\binom{2j}{2i+1}$. For a more detailed argument cf. 4.1. □

6.2. PROOF OF THEOREM 5.2. Denote $\eta_j(s)$ the right-hand side of (5.4). By $\pi_j, \eta_j \in \mathcal{P}_j$, it is enough to prove (5.4) for at least $j+1$ values. Namely we prove $\pi_j(t) = \eta_j(t)$ if $t = 0, 1, \dots, m$ (consider relation $j+1 \leqq \left[\frac{m}{2}\right] + 1 \leqq m$).

We apply induction. If $t = 0$, (5.4) gives $\eta_j(0) = 0$ (if $j = 1$, $\{\dots\} = 0$; when $j > 1$, $p_{j-1}(0) = 0$ (cf. (3.5))), whence using (5.2), (5.4) holds true for $t = 0$. Now we suppose $\pi_j(i) = \eta_j(i)$ if $i = 0, 1, \dots, t$ and prove it for $i = t+1$. By (5.2)

$$\pi_j(t+1) = \sum_{i=0}^{j-1} \binom{2j-1}{2i} \left\{ \frac{\alpha-\beta}{2} + \left(\frac{\alpha+\beta}{2} + j \right) x_k \right\} p_i(t) -$$

$$- \sum_{i=0}^{j-1} \binom{2j-1}{2i} i x_k p_i(t) + \frac{1}{2j} \sum_{i=0}^{j-1} \binom{2j}{2i+1} \times$$

$$\times (2i+1) \left\{ \frac{\alpha-\beta}{2} t + \left(\frac{\alpha+\beta}{2} t + t + i \right) x_k \right\} p_i(t) := S_1 - S_2 + S_3.$$

Here, by (3.3)

$$S_1 = (2j-1) \left\{ \frac{\alpha-\beta}{2} + \left(\frac{\alpha+\beta}{2} + j \right) x_k \right\} p_{j-1}(t+1).$$

Further, by

$$\binom{2j-1}{2i} = \frac{1}{2j}\binom{2j}{2i+1}(2i+1)$$

and again by (3.3)

$$S_3 - S_2 = (2j-1) \left\{ \frac{\alpha-\beta}{2} t + \left(\frac{\alpha+\beta}{2} t + t \right) x_k \right\} p_{j-1}(t+1),$$

whence $S_1 - S_2 + S_3 = \eta_j(t+1)$. □

6.3. PROOF OF COROLLARY 5.3. Let $\beta = -1+\gamma$, $\alpha = \beta + 2\rho$ and $x_k \geqq \varepsilon$ ($\gamma > 0$, $\rho \geqq 0$, $\varepsilon > 0$). If (5.5) were not hold then with a proper subsequence $\mathcal{N}$, $|\pi_j(t, x_{kn})| = \mathbf{o}(1)$ if $n \in \mathcal{N}$, $x_{kn} \in [\varepsilon, 1)$. By (5.4), using $p_{j-1}(t) \geqq B$ (cf. (3.5))

$$\mathbf{o}(1) = \{\dots\}_{s=t} = \rho t + ((-1+\gamma+\rho)t + t + j - 1)x_k, \quad n \in \mathcal{N},$$

whence by $t \geqq 1$, $\varepsilon \leqq x_k < 1$ and $j \geqq 1$

$$\mathbf{o}(1) = \frac{\rho}{x_k} + \gamma + \rho + \frac{j-1}{t} > 2\rho + \gamma, \quad n \in \mathcal{N}, \tag{6.1}$$

a contradiction. Now let $x_{kn} \leqq -\delta_1$. Analogously as above

$$\mathbf{o}(1) = \left|\gamma + \rho + \frac{j-1}{t} + \frac{\rho}{x_k}\right| \geqq \gamma + \rho - \frac{\rho}{\delta_1} > \frac{\gamma}{2}, \quad n \in \mathcal{N}$$

(if $\delta_1 < 1$ is big enough), a contradiction. □

PROOF OF COROLLARY 5.4. If $\alpha = \beta + 2\rho$ (ρ may be negative) and $\beta = -1+\gamma$, $\gamma > 0$, we get supposing that (5.6) does not hold

$$\frac{\rho}{x_k} + \gamma + \rho + \frac{1}{m} - \frac{1}{2} = \mathbf{o}(1), \quad n \in \mathcal{N} \tag{6.2}$$

(cf. (6.1)). By (6.2)

$$\alpha = -1 + \gamma + 2\rho = -\frac{1}{2} - \frac{1}{m} + \rho\left(1 - \frac{1}{x_k}\right) + \mathbf{o}(1), \quad n \in \mathcal{N}$$

which gives a contradiction in both cases (i) and (ii) if $a_1 (\approx 1)$ and $a_2 (\approx 1)$ are properly chosen (when $\alpha \approx -\frac{1}{2} - \frac{1}{m}$). Now we consider (iii). Again from (6.2)

$$\beta = -1 + \gamma = -\frac{1}{2} - \frac{1}{m} - \rho\left(1 + \frac{1}{x_k}\right) + \mathbf{o}(1), \quad n \in \mathcal{N}, \tag{6.3}$$

if (5.6) were not true. But (6.3) contradicts to (iii) with a proper $a_3 (\approx -1)$. Finally, by (iv) $\beta = -\frac{1}{2} - \frac{1}{m} - \delta_5$ ($\delta_5 > 0$), so (6.3) can be used again. □

6.4. PROOF OF THEOREM 5.5. Let

$$\begin{aligned} D_\ell(s) = &\sum_{u=0}^{\ell} (-1)^u \binom{2\ell+1}{2u} R_u(s)\pi_{\ell-u+1}(s) + \\ &+ \sum_{u=0}^{\ell-1} (-1)^u \binom{2\ell+1}{2u+1} Q_{u+1}(s)p_{\ell-u}(s). \end{aligned} \tag{6.4}$$

We prove (cf. (4.12))

$$D_\ell(t)=0, \quad \ell\geqq 0, \quad |t|=0,1,\dots,m. \tag{6.5}$$

If $\ell=0$, (6.5) comes from $\pi_1(s)=-Q_1(s)$ (cf. (5.7) and (5.4)). To go further we prove $D_\ell(-s)=D_\ell(s)$. Indeed, by (3.9) and (5.7)

$$D_\ell(-s)=\sum_{u=0}^{\ell}\binom{2\ell+1}{2n}p_u(s)(-1)^{\ell-n}Q_{\ell-u+1}(s)+$$
$$+\sum_{u=0}^{\ell}\binom{2\ell+1}{2u+1}\pi_{u+1}(s)(-1)^{\ell-u}R_{\ell-u}(s):=I$$

which, applying the new variable $v=\ell-u$, gives $I=D_\ell(s)$. So it is enough to prove (6.5) for $t=0,1,2,\dots,m$ (because D_ℓ is even). Now, induction for t ($\ell\geqq 1$, fixed). If $t=0$, $D_\ell(0)=0$ is obvious by (5.7) and $\pi_j(0)=0$.

Supposing $D_\ell(0)=D_\ell(\pm 1)=\dots=D_\ell(\pm t)=0$ we prove $D_\ell(t+1)=0$. First a remark. Using (5.2), simple computation gives

$$\pi_j(s+1)=\frac{1}{2}\sum_{u=0}^{j-1}\left\{(\alpha-\beta+(\alpha+\beta+1)x_k)\binom{2j-1}{2u}+\right.$$
$$\left.+(2j-1)x_k\binom{2j-2}{2u}\right\}p_u(s)+\frac{1}{2j}\sum_{u=0}^{j-1}\binom{2j}{2u+1}\pi_{u+1}(s), \tag{6.6}$$

so by (6.4), (6.6) and (3.3)

$$0=D_\ell(-t)=\sum_{u=0}^{\ell}(-1)^u\binom{2\ell+1}{2u}R_u(-t)\frac{1}{2}\sum_{\nu=0}^{\ell-u}\left\{(\alpha-\beta+(\alpha+\beta+1)x_k)\times\right.$$
$$\left.\times\binom{2(\ell-u)+1}{2\nu}+(2\ell-2u+1)x_k\binom{2\ell-2u}{2\nu}\right\}p_\nu(-t-1)+$$
$$+\sum_{u=0}^{\ell}(-1)^u\binom{2\ell+1}{2u}R_u(-t)\frac{1}{2\ell-2u+2}\sum_{\nu=0}^{\ell-u}\binom{2\ell-2u+2}{2\nu+1}\pi_{\nu+1}(-t-1)+$$
$$+\sum_{u=0}^{\ell}(-1)\binom{2\ell+1}{2u+1}Q_{u+1}(-t)\frac{1}{2\ell-2u+1}\sum_{\nu=0}^{\ell-u}\binom{2\ell-2u+1}{2\nu}p_\nu(-t-1):=$$
$$:=S_1+S_2+S_3.$$

At S_1 we apply $R_n(-t)=(-1)^n p_n(t)$, $p_\nu(-t-1)=(-1)^\nu R_\nu(t+1)$, some

combinatorical identity and change the order of summations, i.e. we get

$$S_1 = \sum_{\nu=0}^{\ell} \binom{2\ell+1}{2\nu} (-1)^\nu R_\nu(t+1) \frac{1}{2} \sum_{u=0}^{j-\nu} \Big\{ (\alpha - \beta + (\alpha+\beta+1)x_k) \times$$

$$\times \binom{2\ell-2\nu+1}{2u} + (2\ell-2\nu+1)x_k \binom{2\ell-2\nu}{2u} \Big\} p_u(t).$$

Using similar considerations for S_2 and S_3, we get $0 = S_1 + S_2 + S_3 = D_\ell(t+1)$.

Now we can prove (5.8) by induction. Let $E(2t+1)$ denote the right-hand side of (5.8). Then, by $e_{1knm} = -(\ell_k^m)'$ (cf. (2.2) and (3.10)), $E(1) = e_{1knm}$ (cf. (5.7), (5.2) and (2.9)). Suppose $E(2i-1) = e_{2i-1,knm}$, $i = 1, 2, \dots, t$, and we prove it for $i = t+1$. Indeed, by (3.10), (5.1) and (3.1)

$$e_{2t+1,knm} = -\frac{1}{(2t+1)!} \Big\{ \sum_{i=0}^{t} (2t+1)_{2i} e_{2i,k} (\ell_k^m)^{(2t+1-2i)} +$$

$$+ \sum_{i=0}^{t-1} (2t+1)_{2i+1} e_{2i+1,k} (\ell_k^m)^{(2t-2i)} \Big\} = -\frac{1}{(2t+1)!} \Big\{ \sum_{i=0}^{t} \frac{(2t+1)_{2i}}{(2i)!} (1+\varepsilon_k) \times$$

$$\times R_i(m) \left(\frac{n}{\sin \vartheta_k} \right)^{2i} \frac{(-1)^{t-i}}{\sin^2 \vartheta_k} \left(\frac{n}{\sin \vartheta_k} \right)^{2t-2i} (\pi_{t-i+1}(m) + \varepsilon_k) +$$

$$+ \sum_{i=0}^{t-1} \frac{(2t+1)_{2i+1}}{(2i+1)!} \frac{Q_{i+1}(m) + \varepsilon_k}{\sin^2 \vartheta_k} \left(\frac{n}{\sin \vartheta_k} \right)^{2i} (-1)^{t-i} \times$$

$$\times \left(\frac{n}{\sin \vartheta_k} \right)^{2t-2i} p_{t-i}(m)(1+\varepsilon_k) \Big\} =$$

$$= \frac{1}{(2t+1)! \sin^2 \vartheta_k} \left(\frac{n}{\sin \vartheta_k} \right)^{2t} \times$$

$$\times (-1)^{t+1} \Big\{ \sum_{i=0}^{t} \binom{2t+1}{2i} (-1)^i R_i(m) (\pi_{t-i+1}(m) + \varepsilon_k) \times$$

$$\times \sum_{i=0}^{t-1} \binom{2t+1}{2i+1} (-1)^i (Q_{i+1}(m) + \varepsilon_k) p_{t-i}(m) \Big\}.$$

Here, by $D_{t+1}(m) = 0$, we get $\{\dots\} = (-1)^{t+1}(Q_{t+1}(m) + \varepsilon_k)$ whence we get (5.8) for $i = t+1$ (cf 4.3). □

REFERENCES

[1] SAKAI, R. and VÉRTESI, P., Hermite-Fejér interpolations of higher order III, *Studia Sci. Math. Hungar.* **28** (1993), 87–97.

[2] VÉRTESI, P., Hermite–Fejér interpolations of higher order. I, *Acta Math. Hungar.* **54** (1989), 135–152. *MR* **90k:** 41008

[3] VÉRTESI, P., Hermite-Fejér interpolations of higher order. II, *Acta Math. Hungar.* **56** (1990), 369–380.

(Received May 2, 1990)

DEPARTMENT OF MATHEMATICS
KARIYA-HIGASHI SENIOR HIGH SCHOOL
HAJODO-CHO MITSUMATA 20
AICHI KARIYA 448
JAPAN

MTA MATEMATIKAI KUTATÓINTÉZETE
P.O.BOX 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 387–393

ÜBER EINE DARSTELLUNG VON LÖSUNGEN DER KOMPLEXEN DIFFERENTIALGLEICHUNG $w_{\bar{z}} = b_{(F,G)}\overline{w}$

K. KOCA

1. Einleitung

Es sei $D_0 \subset \mathbf{C}$ ein einfach zusammenhängendes Gebiet und $H^1_{D_0}$ die Klasse der über D_0 erklärten komplexwertigen Funktionen, die partielle Ableitungen erster Ordnung nach x und y besitzen und Hölder-stetig mit $z = z + iy \in D_0$ sind, dann nennt man die Menge

$$E_{D_0} := \{E = (F, G) | E : D_0 \to \mathbf{C} \times \mathbf{C},$$

$$\forall z \in D_0; E \in H^1_{D_0} \times H^1_{D_0}, \forall z \in D_0, \operatorname{Im}(\bar{F}G) > 0\}$$

Erzeugendenraum und jedes Element von E_{D_0} heißt Erzeugendenvektor.

Bezeichnet ΩD den Vektorraum über $\mathbf{R}$ der in $D \subset\subset D_0$ reellen Vektorfunktionen $\omega = \begin{pmatrix} \varphi \\ \psi \end{pmatrix}$, so heißt $w = E\omega \in E\Omega D$ in D eine pseudoholomorphe Funktion 1.Art, wenn für $z = x + iy$

$$(1) \qquad \bigwedge_{z_0 \in D} \frac{d_E w}{dz}(z_0) := \dot{w}(z_0) := \lim_{z \to z_0} \left[E(z) \frac{\omega(z) - \omega(z_0)}{z - z_0} \right]$$

eigentlich existiert. Wir wollen die Menge der pseudoholomorphen Funktionen mit $P_D(E)$ bezeichnen. Es ist bekannt, daß die Menge $P_D(E)$ einen additiven Vektorraum über $\mathbf{R}$ bildet, wenn $P_D(E)$ vorhanden ist.

$$\text{Ist } w = E\omega \in P_D(E), \text{ so ist } \omega \in C^1 \times C^1, \ \dot{w} = E\omega_z \text{ und } E\omega_{\bar{z}} = 0.$$

Ist speziell $E = (1, i) =: A$, so erweist sich $A\omega$ als holomorph in D, $P_D(A)$ bezeichne demgemäß die Menge der in D holomorphen Funktionen.

Nach L. Bers [3] nennt man die durch $E = (F, G)$ erklärten Funktionen

---

1980 *Mathematics Subject Classification* (1985 Revision). Primary 30G20; Secondary 30D60.

Key words and phrases. Pseudoholomorphic functions, generating vector, generating series.

$$(2)\qquad \begin{aligned} a_E &:= -(\bar{F}G_{\bar{z}} - F_{\bar{z}}\bar{G})\Delta, & b_E &:= (FG_{\bar{z}} - F_{\bar{z}}G)/\Delta \\ A_E &:= -(\bar{F}G_z - F_z\bar{G})/\Delta, & B_E &:= (FG_z - F_zG)/\Delta \end{aligned}$$

die charakteristischen Koeffizienten mod E, wobei $\Delta = F\bar{G} - \bar{F}G$. Außerdem nennen wir die Funktion

$$(3)\qquad {}_*w = \frac{\bar{G} - i\bar{F}}{\Delta} w - \frac{G - iF}{\Delta}\bar{w} = \varphi + i\psi$$

pseudoholomorphe Funktion 2.Art mod E.

Wenn $w \in P_D(E)$ ist, dann erfüllt w die Differentialgleichung

$$(4)\qquad w_{\bar{z}} = a_E w + b_E \bar{w}.$$

Wenn die Lösungen von (4) die Darstellung $w = F\varphi + G\psi$ mit $\begin{pmatrix}\varphi\\ \psi\end{pmatrix} \in \Omega D$ besitzen, dann ist

$$(5)\qquad \begin{gathered} F\varphi_{\bar{z}} + G\psi_{\bar{z}} = 0, \\ \dot{w} = \frac{d_{(F,G)}w}{dz} = w_z - A_E w - B_E \bar{w} = F\varphi_z + G\psi_z. \end{gathered}$$

Aus der Theorie der pseudoholomorphen Funktionen wissen wir, daß $\dot{w}$ in (1) sowie die Potenzen w^n mit $n \in \mathbf{Z}$ keine Elemente von $P_D(E)$ sind.

2. Eine Darstellung von Lösungen der Differentialgleichung $w_{\bar{z}} = b_{(F,G)}\bar{w}$

In [5] (siehe etwa [4]) ist die Differentialgleichung

$$w_{\bar{z}} = a_{(F,G)} w$$

mit $G = fF$, $f \in P_D(A)$ behandelt worden. Die Differentialgleichung (4) geht durch die Transformation $W = we^A$ unter Bedingung $a_E = A_{\bar{z}}$

$$(6)\qquad W_{\bar{z}} = \alpha \bar{W}$$

über, wobei $\alpha = b_E e^{\bar{A} - A}$ ist. Bezeichne der Koeffizient $\alpha(z, \bar{z})$ in (6) eine differenzierbare Funktion, die im betrachteten Gebiet nicht verschwindet, dann erfüllen die Lösungen von (6) auch die elliptische Differentialgleichung

$$(7)\qquad W_{z\bar{z}} - (\log \alpha)_z W_{\bar{z}} - \alpha\bar{\alpha} W = 0.$$

Die Lösungen von (7) hat I. N. Vekua [7] als komplexe Potentiale der Differentialgleichung (6) genannt und die Lösungen von (6) sind mit Hilfe gewisser

Integraloperatoren dargestellt worden. Andererseits hat man in [2] die Differentialgleichung (6) durch Differentialoperatoren bei den komplexen partiellen Differentialgleichungen behandelt (man vergleiche auch [1]).

In der vorliegenden Arbeit wollen wir die Differentialgleichung

$$w_{\bar{z}} = b_{(F,G)}\bar{w} \tag{8}$$

im Sinne von L. Bers behandeln. Es sei $w \in P_D(E)$ und wir nehmen nun an, daß w der Differentialgleichung

$$w_{\bar{z}} = b_{(F,G)}\bar{w}$$

genügt und F reellwertig ist. Dann ist aus (2) $a_{(F,G)} \equiv 0$, d.h.

$$G_{\bar{z}} = (F_{\bar{z}}/F)\bar{G}. \tag{9}$$

Weil $G \in P_D(E)$ (siehe etwa [3]) ist, muß $b_{(F,G)} = F_{\bar{z}}/F$ sein. Also ist die charakteristische Koeffizient b_E unabhängig von G. Dann gibt es eine Relation zwischen F und G. Wir nehmen an, daß die Funktion G in der Form

$$G(z) = h(z)F^n(z) \tag{10}$$

mit $h \in C^1(D), n \in \mathbf{Z}$ schreibbar ist. Setzt man (10) in (9) ein, dann ergibt sich durch Vergleich der Koeffizienten von F^n und F^{n+1}

$$h_{\bar{z}} = 0 \qquad nh = \bar{h}. \tag{11}$$

Also h ist holomorph in D_0. Aus der zweiten Gleichung in (11) erhält man $n\bar{h} = h$, d.h. $n(h - \bar{h}) = \bar{h} - h$. Daraus folgt $n = -1$. Also ist $h = ih_1$, wobei h_1 reellwertig ist. Weil h holomorph ist, muß $h_1 = c \in R$ konstant sein. Dann ist $E = (F, icF^{-1}) \in E_{D_0}$ mit $c \in R^+$.

Weil die charakteristischen Koeffizienten

$$a_{(F,icF^{-1})} \equiv 0, b_{(F,icF^{-1})} = (\operatorname{Log}|F|)_{\bar{z}}, A_{(F,icF^{-1})} \equiv 0, B_{(F,icF^{-1})} = (\operatorname{Log}|F|)_z$$

sind, ist die Funktion

$$w = F\varphi + icF^{-1}\psi \tag{12}$$

die Lösungen von (8) unter Bedingung

$$F\varphi_{\bar{z}} + icF^{-1}\psi_{\bar{z}} = 0. \tag{13}$$

Andererseits ist (13) equivalent mit dem reellen System

$$\begin{aligned} F^2\varphi_x - c\psi_y &= 0 \\ F^2\varphi_y + c\psi_x &= 0. \end{aligned} \tag{14}$$

Hieraus erhält man

$$\varphi_{xx} + \varphi_{yy} + 2(\mathrm{Log}|F|)_x\varphi_x + 2(\mathrm{Log}|F|)_y\varphi_y = 0. \tag{15}$$

Wenn φ eine reelle Lösung von (15) ist, so kann man die Funktion ψ durch das Kurvenintegral

$$\psi(z) = \int_{z_0}^{z} \left(\frac{-F^2}{c}\varphi_y\right) dx + \left(\frac{F^2}{c}\varphi_x\right) dy$$

berechnen, wobei $z, z_0 \in D$, $z \neq z_0$ ist.

Es sei w eine Lösung von (8) und wir betrachten nun die Form

$$w = \Phi(z)\varphi + \beta(z)\varphi_z \tag{16}$$

wobei φ eine Lösung von (15) und $\Phi, \beta \in H_D^1$, $\Phi \neq 0$, $\beta \neq 0$ in D sind. (Vergleiche etwa [6].) Setzt man die Funktion (16) in (8) ein, so erhält man

$$\beta\varphi_{z\bar{z}} + \left(\Phi - \frac{F_{\bar{z}}}{F}\bar{\beta}\right)\varphi_{\bar{z}} + \beta_{\bar{z}}\varphi_z + \left(\Phi_{\bar{z}} - \frac{F_{\bar{z}}}{F}\bar{\Phi}\right)\varphi = 0. \tag{17}$$

Es sei Φ eine partikuläre Lösung von (8). Dann ergibt sich

$$\varphi_{xx} + \varphi_{yy} + \frac{2}{\beta}\left(\Phi - \frac{F_{\bar{z}}}{F}\bar{\beta} + \beta_{\bar{z}}\right)\varphi_x + \frac{2i}{\beta}\left(\Phi - \frac{F_{\bar{z}}}{F}\bar{\beta} - \beta_{\bar{z}}\right)\varphi_y = 0. \tag{18}$$

Weil φ eine Lösung von (15) ist, erhält man aus (18)

$$\begin{aligned} \Phi - (\mathrm{Log}|F|)_{\bar{z}}\bar{\beta} + \beta_{\bar{z}} &= (\mathrm{Log}|F|)_x\beta \\ \Phi - (\mathrm{Log}|F|)_{\bar{z}}\bar{\beta} - \beta_{\bar{z}} &= -i(\mathrm{Log}|F|)_y\beta. \end{aligned} \tag{19}$$

Also ist die Funktion β in der Form $\beta = gF$ schreibbar, wobei $g \in P_D(A)$ ist. Aus (19) erhält man

$$\Phi = \beta(F_z/F) + \bar{\beta}(F_{\bar{z}}/F) = gF_z + \bar{g}F_{\bar{z}}.$$

So ist Φ eine reellwertige Lösung von (8). Dann muß Φ die Differentialgleichung

$$\Phi_{\bar{z}} = b_{(F,icF^{-1})}\Phi \tag{20}$$

erfüllen. Nach [5] ist Φ in der Form $\Phi = \sigma F$ mit $\sigma \in P_D(A)$ schreibbar. Weil Φ reellwertig ist, muß die Funktion σ eine reelle Konstant sein. Hieraus erhält man $\Phi = kF$ mit $k \in \mathbf{R}$. Setzt man Φ und β in (16) ein, dann ist

$$w = F(k\varphi + g\varphi_z). \tag{21}$$

Außerdem gilt die Relation $\mathrm{Re}(gF_z) = \frac{k}{2}F$.

FOLGERUNG 1. *Wenn die Funktion φ eine reellwertige Lösung von* (15) *ist, dann genügt* (21) *der Differentialgleichung*

$$w_{\bar{z}} = b_{(F,icF^{-1})}\bar{w}$$

unter Bedingungen (19) *mit* $g = F^{-1}\beta$.

SATZ 1. *Wenn* $w \in P_D(F, icF^{-1})$ *ist, dann existiert* $w_1 \in P_D(F^2, ic)$, *sodaß* $w_1 = Fw$ *ist und*

$$(w_1)_{\bar{z}} = F_{\bar{z}}(w + \bar{w}) = a_{(F^2,ic)}w_1 + b_{(F^2,ic)}\bar{w}_1$$

$$\dot{w}_1 = \frac{d_{(F^2,ic)}w_1}{dz} = F\frac{d_{(F,icF^{-1})}w}{dz}.$$

BEWEIS. Der Beweis folgt unmittelbar aus der Definition von $\dot{w}$ und unter Berücksichtigung von (12).

FOLGERUNG 2. $P_D(F, icF^{-1}) = \frac{1}{F}P_D(F^2, ic)$.

FOLGERUNG 3. *Die Erzeugendenvektoren* $E := (F, icF^{-1})$ *und* $E^* := (F^2, ic)$ *sind ähnlich in* D_0. (Siehe etwa [3], Seite 38.)

Wir definieren zwei Funktionen durch $E = (F, G) \in E_{D_0}$:

$$P_{(F,G)} := \frac{\bar{G} - i\bar{F}}{\Delta}; \qquad Q_{(F,G)} := -\frac{G - iF}{\Delta}.$$

FOLGERUNG 4. *Es seien* $w_1 \in P_D(F^2, ic)$, $w \in P_D(F, icF^{-1})$. *So ist*

$$P_{(F,icF^{-1})}w_1 + Q_{(F,icF^{-1})}\bar{w}_1 = F\omega,$$

$$P_{(F,icF^{-1})} = FP_{(F^2,ic)}, \qquad Q_{(F,icF^{-1})} = FQ_{(F^2,ic)}$$

wobei $\omega = \varphi + i\psi$ (F, icF^{-1})*-pseudoholomorph 2.Art ist.*

FOLGERUNG 5. *Wenn* $w \in P_D(F, icF^{-1})$ *ist, so ist*

$$P_{(F^2,ic)}w + Q_{(F^2,ic)}\bar{w} = \omega_1/F,$$

wobei die Funktion ω_1 (F^2, ic)*-pseudoholomorph 2.Art ist.*

FOLGERUNG 6.

$$\frac{\partial}{\partial\bar{z}}P_{(F,icF^{-1})} = -b_{(F,icF^{-1})}\overline{Q_{(F,icF^{-1})}}$$

$$\frac{\partial}{\partial\bar{z}}Q_{(F,icF^{-1})} = -b_{(F,icF^{-1})}\overline{P_{(F,icF^{-1})}}.$$

Es seien $E = (F, icF^{-1})$, $E_n := (F^n, icF^{-n})$ mit $n \in \mathbf{Z}$. Dann sind die charakteristische Koeffizienten

$$a_{E_n} \equiv 0, \quad b_{E_n} = nb_E, \quad A_{E_n} \equiv 0, \quad B_{E_n} = nB_E.$$

SATZ 2. *Wenn $w \in P_D(F, icF^{-1})$ und w eine reellwertige Lösung von* (8) *ist, so ist $w^n \in P_D(E_n)$ und*

$$(w^n)_{\bar{z}} = b_{E_n} w^n,$$

$$\frac{d_{(F^n, icF^{-n})} w^n}{dz} = nw^{n-1} \frac{d_E w}{dz}.$$

BEWEIS. Der Beweis folgt aus (4) und (5).

SATZ 3. *Wenn w_n^* eine Lösung der Differentialgleichung*

$$(w_n^*)_{\bar{z}} = (-1)^n \left(\frac{h}{\bar{h}}\right)^n b_{E_n} \bar{w}_n^*$$

ist, so ist $w_n^* \in P_D(E_n^*)$, wobei $n \in \mathbf{Z}$, $h \in P_D(A)$, $E_n^* := [(ihF)^n, i(ichF^{-1})^n] \in \in E_{D_0}$.

BEWEIS. Der Beweis folgt unmittelbar aus der Definition von $P_D(F, G)$.

FOLGERUNG 7. *Wenn $h = F_z$, d.h. $F_{z\bar{z}} = 0$ ist, dann ist $\tilde{E}_\nu := [(ih)^\nu F, i(ich)^\nu F^{-1}]$ mit $\nu \in \mathbf{Z}$ eine Erzeugendenfolge im Bers'schen Sinne für die Differentialgleichung* (8) (siehe etwa [2]).

REFERENCES

[1] BAUER, K. W., Zur Darstellung pseudoanalytischer Funktionen, *Function theoretic methods for partial differential equations* (Proc. Internat. Sympos., Darmstadt, 1976), Lecture Notes in Math., Vol. 561, Springer–Verlag, Berlin, 1976, 101–111. *MR* **57**#6459

[2] BAUER, K. W. und RUSCHEWEYH, S., Ein Darstellungssatz für eine Klasse pseudoanalytischer Funktionen, *Three articles on pseudoanalytic functions and elliptic differential equations*, Gesellsch. Math. Datenverarbeitung, Bonn, Ber. No. 75, Gesellsch. Math. Datenverarbeitung, Bonn, 1973, 3–15. *MR* **54**#10618

[3] BERS, L., *Theory of pseudo-analytic functions*, Lecture Notes, Institute for Mathematics and Mechanics, New York University, New York, 1953. *MR* **15**–211

[4] KOCA, K., Über die Eigenschaften der Räume $P_D(F^n, fF^n)$ und $P_D(F^n, iF^n)$, *Facta Univ. Ser. Math. Inform.* **3** (1988), 13–21. *MR* **91b**:30138

[5] KOCA, K. und WITHALM, C., Über die Eigenschaften des Raumes $P_D(F, fF)$, *An. Ştiinţ. Univ. „Al. I. Cuza" Iaşi Secţ. Ia Mat. (N. S.)* **35** (1989), 101–107. *MR* **91c**:30091

[6] TUTSCHKE, W., Konstruktion von Lösungen gewisser spezieller elliptischer Differentialgleichungen zweiter Ordnung (in der Ebene) durch Zurückführung auf die inhomogene Cauchy–Riemannsche Differentialgleichung, *Math. Nachr.* **82** (1978), 69–75. *MR* **58**#6664

[7] VEKUA, I. N., *Verallgemeinerte analytische Funktionen*, Akademie-Verlag, Berlin, 1963. *MR* **28**#1312

(Eingegangen am 12. Mai 1990.)

ANKARA ÜNIVERSITESI FEN FAKÜLTESI
MATEMATIK BÖLÜMÜ
TANDOĞAN
TR–06100 ANKARA
TURKEY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 395–400

A NOTE ON SUMMABILITY

M. A. SARIGÖL

Abstract

In this paper we give necessary and sufficient conditions in order that the summability $|\bar{N}, p_n| \Rightarrow |\bar{N}, q_n|_k, k \geq 1$. Therefore we extend the known results of [1], [4] to the case $k > 1$. Moreover we discuss its special cases.

1. Introduction

Let $\sum a_n$ be a given infinite series with s_n as its nth partial sum. If (p_n) is a sequence of positive constants, and

$$P_n = p_0 + p_1 + \ldots + p_n \to \infty \quad \text{as } n \to \infty \qquad (P_{-v} = p_{-v} = 0, v \geq 1)$$

then the Riesz mean t_n of $\sum a_n$ is defined by

$$t_n = \frac{1}{P_n} \sum_{v=0}^{n} p_v s_v, \qquad (P_n \neq 0).$$

If $(t_n) \in bv$, i.e.,

$$\sum_{n=0}^{\infty} |t_n - t_{n-1}| < \infty, \qquad (t_{-1} = 0)$$

then the series $\sum a_n$ is said to be summable $|\bar{N}, p_n|$. Concerning the $|\bar{N}, p_n|$-summability the following result is due to Sunouchi [1].

THEOREM 1.1. *Let (p_n) and (q_n) be positive sequences such that*

$$q_n P_n = O(p_n Q_n) \quad \text{as } n \to \infty. \tag{1}$$

Then $|\bar{N}, p_n| \Rightarrow |\bar{N}, q_n|$.

In 1950, while reviewing this paper, Bosanquet [4] observed that Condition (1) is not only sufficient but also necessary for $|\bar{N}, p_n| \Rightarrow |\bar{N}, q_n|$. Later

1980 *Mathematics Subject Classification* (1985 Revision). Primary 40G99; Secondary 40F05.

Key words and phrases. Absolute and strong summability, direct theorem on summability.

on, the concept of the summability $|\bar{N}, p_n|$ was extended to the concept of the summability $|\bar{N}, p_n|_k$, $k \geq 1$, defined by

$$\sum_{n=1}^{\infty} \left(\frac{P_n}{p_n}\right)^{k-1} |t_n - t_{n-1}|^k < \infty, \qquad [2].$$

We may also note that, while it is clear that the $|\bar{N}, p_n|_k$-summability with $k = 1$ reduces to the $|\bar{N}, p_n|$-summability, these methods are in general independent of each other, for $k > 1$. Therefore this also raises the problem: what are the necessary and sufficient conditions for $|\bar{N}, p_n| \Rightarrow |\bar{N}, p_n|_k$, $k > 1$ and conversely. We obtain an affirmative answer to this question.

We require the following lemma and theorems.

LEMMA 1.2. *Suppose that $k > 0, p_n > 0, P_n \to \infty$ as $n \to \infty$. Then there exist two (strictly) positive constants M and N, depending only k, for which for all $v \geq 1$,*

$$\frac{M}{P_{v-1}^k} \leq \sum_{n=v}^{\infty} \frac{p_n}{P_n P_{n-1}^k} \leq \frac{N}{P_{v-1}^k}$$

where M and N are independent of (p_n), [2].

THEOREM 1.3. *$T = (a_{nv}) \in (\ell_1, \ell_k)$ if and only if*

$$\sup_v \sum_{n=1}^{\infty} |a_{nv}|^k < \infty \tag{2}$$

for the cases $1 \leq k < \infty$, where (ℓ_1, ℓ_k) denotes the set of all infinite matrices T which map ℓ_1 into $\ell_k = \{a = (a_v) : \sum |a_v|^k < \infty,\ k \geq 1\}$, [3].

THEOREM 1.4. *$T = (a_{nv}) \in (\ell_q, \ell_1)$ if and only if*

$$\sup_E \sum_{v=1}^{\infty} \left| \sum_{n \in E} a_{nv} \right|^q < \infty \tag{3}$$

where $q = k(k-1)^{-1} > 1$ and E is any finite subset of positive integers [5].

2

The purpose of this paper is to derive necessary and sufficient conditions in order that $|\bar{N}, p_n| \Rightarrow |\bar{N}, q_n|_k$, $k \geq 1$, so to extend results [1], [4], and to consider its special cases.

THEOREM 2.1. *A necessary and sufficient condition in order that* $|\bar{N}, p_n| \Rightarrow |\bar{N}, q_n|_k$, $k \geqq 1$, *is*

$$q_v P_v^k = O(p_v^k Q_v) \quad \text{as } v \to \infty. \tag{4}$$

PROOF. Let t_n denote the $(\bar{N}, p_n)$-mean of the series $\sum a_n$. Then, by the definition we have

$$t_n = \frac{1}{P_n} \sum_{v=0}^{n} p_v s_v = \frac{1}{P_n} \sum_{v=0}^{n} (P_n - P_{v-1}) a_v.$$

If $\sum a_n$ is summable $|\bar{N}, p_n|$, then

$$\sum_{n=1}^{\infty} |\Delta t_{n-1}| < \infty. \tag{5}$$

Since

$$\Delta t_{n-1} = \left\{ -\frac{1}{P_{n-1}} + \frac{1}{P_n} \right\} \sum_{v=0}^{n} P_{v-1} a_v$$

$$= -\frac{p_n}{P_n P_{n-1}} \sum_{v=1}^{n} P_{v-1} a_v, \qquad n \geqq 1, \ (P_{-1} = 0)$$

we have, for $n \geqq 1$,

$$P_{n-1} a_n = -\frac{P_n P_{n-1}}{p_n} \Delta t_{n-1} + \frac{P_{n-1} P_{n-2}}{p_{n-1}} \Delta t_{n-2},$$

i.e.,

$$a_n = -\frac{P_n}{p_n} \Delta t_{n-1} + \frac{P_{n-2}}{p_{n-1}} \Delta t_{n-2}, \quad (t_{-1} = 0). \tag{6}$$

If T_n denotes the $(\bar{N}, q_n)$-mean of $\sum a_n$, we get similarly, by (6),

$$\Delta T_{n-1} = -\frac{q_n}{Q_n Q_{n-1}} \sum_{v=1} Q_{v-1} a_v =$$

$$= -\frac{q_n}{Q_n Q_{n-1}} \sum_{v=1}^{n} Q_{v-1} \left\{ -\frac{P_v}{p_v} \Delta t_{v-1} + \frac{P_{v-2}}{p_{v-1}} \Delta t_{v-2} \right\} =$$

$$= \frac{q_n P_n}{p_n Q_n} \Delta t_{n-1} + \frac{q_n}{Q_n Q_{n-1}} \sum_{v=1}^{n-1} Q_{v-1} \frac{P_v}{p_v} \Delta t_{v-1}$$

$$-\frac{q_n}{Q_nQ_{n-1}}\sum_{v=1}^{n-1}Q_v\frac{P_{v-1}}{p_v}\Delta t_{v-1}$$

$$=\frac{q_nP_n}{p_nQ_n}\Delta t_{n-1}+\frac{q_n}{Q_nQ_{n-1}}\sum_{v=1}^{n-1}(Q_{v-1}P_v-Q_vP_{v-1})\frac{\Delta t_{v-1}}{p_v}$$

$$=\frac{q_nP_n}{p_nQ_n}\Delta t_{n-1}+\frac{q_n}{Q_nQ_{n-1}}\sum_{v=1}^{n-1}\left(Q_v-\frac{P_v}{p_v}q_v\right)\Delta t_{v-1}. \tag{7}$$

Now, if we write $\alpha_n=\Delta t_{n-1}$ and $\beta_n=\left(\frac{Q_n}{q_n}\right)^{1-k^{-1}}\Delta T_{n-1}$ for all $n\geq 1$, then

$$\beta_n=\frac{q_nP_n}{Q_np_n}\left(\frac{Q_n}{q_n}\right)^{1-k^{-1}}\alpha_n+\frac{q_n}{Q_nQ_{n-1}}\left(\frac{Q_n}{q_n}\right)^{1-k^{-1}}\sum_{v=1}^{n-1}\left(Q_v-\frac{P_v}{p_v}q_v\right)\alpha_v,$$

i.e., (β_n) is $T=(a_{nv})$ transform sequence of $(\alpha_n)\in\ell_1$, where

$$a_{nv}=\begin{cases}\dfrac{q_n}{Q_nQ_{n-1}}\left(\dfrac{Q_n}{q_n}\right)^{1-k^{-1}}\left(Q_v-\dfrac{P_v}{p_v}q_v\right), & \text{if } 1\leq v\leq n-1\\ \dfrac{q_nP_n}{p_nQ_n}\left(\dfrac{Q_n}{q_n}\right)^{1-k^{-1}}, & \text{if } v=n\\ 0, & \text{if } v>n.\end{cases}$$

So, whenever $\sum a_n$ is summable $|\bar{N},p_n|$, i.e., $(\alpha_n)\in\ell_1$, it is also summable $|\bar{N},q_n|_k$, $k\geq 1$, that is, $(\beta_n)\in\ell_k$ if and only if $T\in(\ell_1,\ell_k)$. On the other hand,

$$\sum_{n=1}^{\infty}|a_{nv}|^k=\left(\frac{q_vP_v}{p_vQ_v}\right)^k\left(\frac{Q_v}{q_v}\right)^{k-1}+\left|Q_v-\frac{P_v}{p_v}q_v\right|^k\sum_{n=v+1}^{\infty}\left(\frac{q_n}{Q_nQ_{n-1}}\right)^k\left(\frac{Q_n}{q_n}\right)^{k-1}.$$

However, the boundedness of the first term on the right-hand side of the last equality implies the boundedness of the second. This can be shown as follows. Since, for $k\geq 1$,

$$\left(\frac{q_v}{Q_v}\right)^k\left(\frac{P_v}{p_v}\right)^k\left(\frac{Q_v}{q_v}\right)^{k-1}=O(1)\Rightarrow\frac{q_v}{Q_v}\frac{P_v}{p_v}=O(1)$$

so by Lemma 1.2,

$$\sum_{n=v+1}^{\infty}\left(\frac{q_n}{Q_nQ_{n-1}}\right)^k\left(\frac{Q_n}{q_n}\right)^{k-1}=O\left(\frac{1}{Q_v^k}\right)\quad\text{as } v\to\infty.$$

It follows that

$$\left|Q_v - \frac{P_v}{p_v}q_v\right|^k \sum_{n=v+1}^{\infty}\left(\frac{q_n}{Q_nQ_{n-1}}\right)^k\left(\frac{Q_n}{q_n}\right)^k = O\left\{1+\frac{q_vP_v}{Q_vp_v}\right\} = O(1).$$

Hence, $\sup_v \sum_{n=1}^{\infty} |a_{nv}|^k < \infty$ is equivalent to Condition (4), which completes the proof of the theorem together with Theorem 1.3.

As it is clearly shown, we get immediately Theorem 1.1 and Bosanquet's result by taking $k=1$ in our above theorem. We can now discuss the other special cases.

THEOREM 2.2. *The necessary and sufficient condition for* $|\bar{N}, p_n| \Rightarrow |\bar{N}, p_n|_k$, $k>1$, *is*

$$P_n = O(p_n) \quad \text{as } n \to \infty. \tag{8}$$

We note that, if we choose $p_n = 1$ for all n, then $P_n = n+1$. In the case, $|\bar{N}, p_n|_k$ reduces to $|C,1|_k$. Since $(n+1) \neq O(1)$, Condition (8) does not hold. Therefore we can derive the following result from Theorem 2.2.

COROLLARY 2.3. *For* $k>1$, $|C,1| \not\Rightarrow |C,1|_k$, *i.e., there exists at least one series that is summable* $|C,1|$ *but not summable* $|C,1|_k$.

COROLLARY 2.4. *The necessary and sufficient condition for* $|\bar{N}, p_n| \Rightarrow |C,1|_k$, $k \geqq 1$, *is*

$$P_n^k = O(np_n^k) \quad \text{as } n \to \infty.$$

COROLLARY 2.5. *For* $k>1$, $|C,1| \not\Rightarrow |\bar{N}, q_n|_k$.

In this case, Condition (4) contradicts to $\sum_{n=1}^{\infty} \frac{q_n}{Q_n} = \infty$.

It is now natural to ask such question as under what necessary and sufficient condition does $|\bar{N}, p_n|_k \Rightarrow |\bar{N}, q_n|$ for $k>1$. But we have not been able to solve this. It is, however, possible to answer the problem in particular case, i.e., $|\bar{N}, p_n|_k \Rightarrow |\bar{N}, p_n|$, $k>1$.

THEOREM 2.6. $|\bar{N}, p_n|_k \not\Rightarrow |\bar{N}, p_n|$ *for* $k>1$.

PROOF. Follow the way in the proof of Theorem 2.1. If we say $\alpha_n = (\frac{P_n}{P_n})^{1-k^{-1}} \Delta t_{n-1}$ and $\beta_n = \Delta t_{n-1}$ for $n \geqq 1$, then by (7) we have

$$\beta_n = \left(\frac{p_n}{P_n}\right)^{1-k^{-1}} \alpha_n,$$

that is, (β_n) is T-transform sequence of the sequence $(\alpha_n) \in \ell_k$, where

$$a_{nv} = \begin{cases} \left(\dfrac{p_n}{P_n}\right)^{1-k^{-1}} & \text{if } v = n \\ 0 & \text{if } v \neq n. \end{cases}$$

On the other hand, for the above matrix, Condition (3) of Theorem 1.4 is reduced to

$$\sum_{n=1}^{\infty} \frac{p_n}{P_n} < \infty,$$

which is not possible. Therefore, $T \notin (\ell_k, \ell_1)$, that is, there exists a series which is summable $|\bar{N}, p_n|_k$ but not summable $|\bar{N}, p_n|$, for $k > 1$.

Considering the above theorem with $p_n = 1$ and Corollary 2.3, the following result can be obtained.

COROLLARY 2.7. *The summabilities* $|C, 1|$ *and* $|C, 1|_k$ *are not equivalent for* $k > 1$.

REFERENCES

[1] SUNOUCHI, G., Notes on Fourier analysis, XVIII. Absolute summability of series with constant terms, *Tôhoku Math. J.* (2) **1** (1949), 57–65. *MR* **11**-654

[2] BOR, H. and KUTTNER, B., On the necessary conditions for absolute weighted arithmetic mean summability factors, *Acta Math. Hungar.* **54** (1989), 57–61. *MR* **90i**:40019

[3] MADDOX, I. J., *Elements of functional analysis,* Cambridge University Press, London–New York, 1970. *MR* **52**#11515

[4] BOSANQUET, L. S., Review on G. Sunouchi's paper "Notes on Fourier analysis. XVIII. Absolute summability of series with constant terms", *Math. Reviews* **11** (1950), 654.

[5] STIEGLITZ, M. and TIETZ, H., Matrixtransformationen von Folgenräumen. Eine Ergebnisübersicht, *Math. Z.* **154** (1977), 1–16. *MR* **56**#5109

(Received May 15, 1990)

ERCIYES ÜNIVERSITESI
FEN-EDEBIYAT FAKÜLTESI
TR–38039 KAYSERI
TURKEY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 401–415

A MESH-INDEPENDENCE PRINCIPLE FOR NONLINEAR OPERATOR EQUATIONS IN BANACH SPACE AND THEIR DISCRETIZATIONS

I. K. ARGYROS

Abstract

The mesh-independence principle states that, when Newton's method is applied to a nonlinear equation between two Banach spaces as well as to some finite-dimensional discretizations of that equation then the number of steps required by the two processes to converge to within a given tolerance is essentially the same. This result has been proved in [2] under the assumption that the Fréchet derivative of the operator is Lipschitz continuous. Here we extend these results to include the case when the derivative of the operator is only Hölder continuous.

Introduction

Consider the equation

$$F(x) = 0 \tag{1}$$

where F is a nonlinear operator defined between two Banach spaces E_1, E. The Newton's method

$$x_{n+1} = x_n - F'(x_n)^{-1} F(x_n), \qquad n = 0, 1, 2, \ldots \tag{2}$$

has been used extensively to approximate a solution x^* of (1). The iterates $\{x_n\}$, $n = 0, 1, \ldots$ can rarely be computed in infinite dimensional spaces.

That is why we replace (1) by a family of discretized equations

$$F_h(z) = 0, \qquad h > 0 \tag{3}$$

where F_h is a nonlinear operator between two finite dimensional spaces E_h^1 and E_h. The discretization on E_1 is defined by the linear operators $L_h : E_1 \to \to E_h^1$.

The Newton's iteration for (3) is given by

$$z_0^h = L_h(x_0), \quad z_{n+1}^h = z_n^h - F_h'(z_n^h)^{-1} F_h(z_n^h), \quad n = 0, 1, \ldots . \tag{4}$$

1980 *Mathematics Subject Classification* (1985 Revision). Primary 65J15; Secondary 65B05, 65L50, 65M50.

Key words and phrases. Newton's method, mesh independence, Hölder continuity.

Akadémiai Kiadó, Budapest

In the excellent paper in reference [2], it is shown that under certain assumptions the solution z_h^* and the iterates z_n^h satisfy the relations

$$z_h^* = L_h(x^*) + O(h^q),$$
$$z_n^h - z_h^* = L_n(x_n - x^*) + O(h^q),$$
$$z_{n+1}^h - z_n^h = L_h(x_{n+1} - x_n) + O(h^q), \qquad q > 0,$$

and for any $\varepsilon > 0$

$$|\min\{n \geqq 0, \|x_n - x^*\| < \varepsilon\} - \min\{n \geqq 0, \|z_n^h - z_h^*\| < \varepsilon\}| \leqq 1$$

for h sufficiently small and x_0 in a ball centered at x^* and of some specific radius $r > 0$.

One of the basic assumptions in [2] is that the Fréchet derivative of F is Lipschitz continuous on a subset $E_2 \subset E_1$.

Here we show that the above results can be extended to include the case when the Fréchet-derivative of F is only (γ, λ)-Hölder continuous (to be precised later) for some $\gamma > 0$ and $\lambda \in [0, 1]$. Our results reduce to the ones in [2] for $\lambda = 1$.

An example is also provided for $\lambda = \frac{1}{2}$ for a scalar, second order, two-point boundary value problem, where our results apply where the ones in [2] do not.

Relevant work has been done in [1], [3], [5]–[7], [10]–[12], and the references there.

To make the paper as self-contained as possible we will use some of the techniques developed in the proofs of the results in [2].

The norms in all spaces will be denoted by the same symbol $\|\ \|$.

DEFINITION. We say that the Fréchet-derivative $F'(x)$ of F is (γ, λ)-Hölder continuous on $E_2 \subset E_1$ if for some $\gamma > 0$, $\lambda \in [0, 1]$

$$\|F'(x) - F'(y)\| \leqq \gamma \|x - y\|^\lambda \quad \text{for all} \quad x, y \in E_2. \tag{5}$$

We then say that $F'(\cdot) \in H_{E_2}(\gamma, \lambda)$.

It is well-known that if E_2 is convex then

$$\|F(x) - F(y) - F'(x)(x - y)\| \leqq \frac{\gamma}{1 + \lambda} \|x - y\|^{1+\lambda} \quad \text{for all} \quad x, y \in E_2. \tag{6}$$

We assume that (1) has a solution $x^* \in E_2$ which is simple in the sense that $F'(x^*)$ has a bounded inverse with norm $\beta = \|F'(x^*)^{-1}\|$.

Finally, we denote by $U(x, r)$, $\overline{U}(x, r)$ the open and closed balls, respectively, with center x and radius $r > 0$.

The following result has been proved in [11] and improved in [10] for $\lambda = 1$.

THEOREM 1. *Let* $F: E_1 \to E$. *Assume* $F'(\cdot) \in H_{E_2}(\gamma, \lambda)$ *on a convex set* $E_2 \subset E_1$. *If* $x^* \in E_2$ *is a solution of*

$$F(x) = 0$$

for which $F'(x^*)$ *is nonsingular, set*

$$U^* = \overline{U}(x^*, r^*),$$

with

$$0 < r^* < \left[\frac{1+\lambda}{(2+\lambda)\beta\gamma}\right]^{1/\lambda}, \quad \lambda \in (0,1] \tag{7}$$

such that $U^* \subset E_2$.

Then, for any $x_0 \in U^*$, *Newton's iteration* (2) *converges to* x^* *and the iterates satisfy*

$$\|x_{n+1} - x^*\| \leqq \frac{\beta\gamma}{1+\lambda} \frac{\|x_n - x^*\|^{1+\lambda}}{1 - \beta\gamma\|x_n - x^*\|^{\lambda}}, \quad n = 0, 1, \ldots . \tag{8}$$

PROOF. By the standard perturbation lemma it follows that $F'(x)$ is nonsingular in U^* and

$$\|F'(x)^{-1}\| \leqq \frac{\beta}{1 - \beta\gamma\|x - x^*\|^{\lambda}}, \quad \text{for all} \quad x \in U^*. \tag{9}$$

Hence Newton's iteration function

$$P(x) = x - F'(x)^{-1}F(x), \quad x \in U^*$$

is well defined on U^* and from

$$\|P(x) - x^*\| \leqq \|F'(x)^{-1}\| \|F(x^*) - F(x) - F'(x)(x^* - x)\| \leqq$$
$$\leqq \frac{\beta}{1 - \beta\gamma\|x - x^*\|^{\lambda}} \frac{1}{\lambda + 1} \gamma \|x - x^*\|^{1+\lambda} \leqq a(r)\|x - x^*\|,$$

for all $x \in U^*$ and

$$a(r) = \frac{\beta\gamma(r^*)^{\lambda}}{(\lambda+1)(1 - \beta\gamma(r^*)^{\lambda})} < 1$$

we obtain the results.

We now state a theorem that can be found in [8, p. 145], whose proof follows exactly as the proof of the Newton–Kantorovich theorem for $\lambda = 1$ [8, p. 143].

THEOREM 2. *Let* $F: E_1 \to E$. *Assume:*

(a) *the linear operator* $F'(\cdot) \in H_{E_3^*}(\gamma, \lambda)$, *where* $E_3^* = U(x_0, R) \subset E_1$

for some $x_0 \in E_1$ *and* $R > 0$;

(b) *the linear operator* $F'(x_0)^{-1}$ *exists and satisfies*

$$\|F'(x_0)^{-1}\| \leqq b_0, \quad \|F'(x_0)^{-1}F(x_0)\| \leqq \eta_0, \quad \ell_0 = b_0\gamma\eta_0^{\lambda} \leqq s \tag{10}$$

where s *is the minimum positive root of the equation*

$$\left(\frac{s}{1+\lambda}\right)^{\lambda} = (1-s)^{1+\lambda} \quad \text{in} \quad \left(0, \frac{1}{2}\right) \quad \text{with} \quad 0 < \lambda < 1. \tag{11}$$

If

$$R \geqq r_0 = \frac{\eta_0}{1-p_0}, \quad \text{where} \quad p_0 = \frac{\ell_0}{(1+\lambda)(1-\ell_0)} \tag{12}$$

then Newton's iteration (2) *converges to a unique solution* x^* *of the equation*

$$F(x) = 0$$

in $\overline{U}(x_0, r_0)$.

As in [2] consider a subset $W^* \subset E_1$ such that

$$x^* \in W^*, \ x_n \in W^*, \ x_n - x^* \in W^*, \ x_{n+1} - x_n \in W^*, \ n = 0, 1, 2, \ldots. \tag{13}$$

Consider the discretization method given by the family

$$\{F_h, L_h, \overline{L}_h\}, \quad h > 0 \tag{14}$$

where

$$F_h \colon D_h \subset E_h^1 \to E_h, \quad h > 0$$

are nonlinear operators and

$$L_h : E_1 \to E_h^1, \ \overline{L}_h \colon E \to E_h, \quad h > 0,$$

are bounded linear discretization operators such that

$$L_h(W^* \cap U^*) \subset D_h, \quad h > 0. \tag{15}$$

The discretization (14) is called λ-*Hölder uniform* if there exist constants $w > 0$, $\ell > 0$ such that

$$\overline{U}(L_h(x^*), w) \subset D_h, \quad h > 0 \tag{16}$$

and

$$\begin{gathered} \|F_h'(w_1) - F_h'(w_2)\| \leqq \ell\|w_1 - w_2\|^{\lambda}, \\ \lambda \in [0, 1), \ h > 0, \ w_1, w_2 \in \overline{U}(L_h(x^*), w). \end{gathered} \tag{17}$$

Moreover, the discretization (14) is called *bounded* if there is a constant $b > 0$ such that

$$\|L_h(u)\| \leqq b\|u\|, \quad u \in W^*, \ h > 0, \tag{18}$$

stable if there is a constant $d > 0$ such that

$$\|F_h'(L_h(u))^{-1}\| \leqq d, \quad u \in W^* \cap U^*, \ h > 0, \tag{19}$$

consistent of order $q > 0$ if there are two constants $c_0 > 0$, $c_1 > 0$ such that

$$\|\overline{L}_h(F(x)) - F_h(L_h(x))\| \leqq c_0 h^q, \quad x \in W^* \cap U^*, \ h > 0 \tag{20}$$

$$\|\overline{L}_h(F'(u))(v) - F_h'(L_h(u))L_h(v)\| \leqq c_1 h^q, \ u \in W^* \cap U^*, \ v \in W^*, \ h > 0. \tag{21}$$

We can now prove the main result:

THEOREM 3. *Let $F: E_2 \subset E_1 \to E$ be an operator satisfying the hypotheses of Theorem 1 and consider a uniform discretization* (14) *which is bounded, stable and consistent of order q. Then* (3) *has a locally unique solution*

$$z_h^* = L_h(x^*) + O(h^q) \tag{22}$$

for all $h > 0$ satisfying

$$0 < h \leqq h_0 = \min\left[\left(\frac{e}{d^2 \ell c_0}\right)^{1/q}, \left(\frac{w}{m d c_0}\right)^{1/q}\right] \tag{23}$$

with $e = \frac{1}{2}$ and $m = \frac{(1+\lambda)(1-e)}{(1+\lambda)(1-e)-e}$.

Moreover, if the following condition is satisfied:

$$T \equiv A\left(\frac{B-C}{2A}\right)^{\lambda+1} + C\left(\frac{B-C}{2A}\right)^{\lambda} - B\left(\frac{B-C}{2A}\right) < 0, \quad C < B \tag{24}$$

with

$$\begin{aligned} A &= (\lambda+2)d\ell \\ B &= \lambda + 1, \ \lambda \in (0,1) \\ C &= 2r^*(\lambda+1)\ell b \\ D &= 2(\lambda+1)c, \quad c = \max(c_0, c_1). \end{aligned}$$

Then there exist constants $h_1 \in (0, h_0]$, $r_1 \in (0, r^*]$ such that Newton's iteration (4) *converges to z_h^* and that*

$$z_n^h = L_h(x_n) + O(h^q), \quad n = 0, 1, 2, \ldots \tag{25}$$

$$z_n^h - z_h^* = L_h(x_n - x^*) + O(h^q), \quad n = 0, 1, 2, \ldots \tag{26}$$

for all $h \in (0, h_1]$, *and all starting points* $z_0 \in \overline{U}(z^*, r_1)$.

PROOF. For simplicity, we will prove the theorem for $\lambda \in (0,1)$. By Theorem 2, when

$$\ell_0 = \ell_0(h) = d\ell \| F_h'(L_h(x^*))^{-1} F_h((L_h(x^*))) \| \leqq s = s(h) < e \tag{27}$$

$$r_0 = r_0(h) = \frac{\eta_0(h)}{1 - p_0(h)} \leqq w,$$

with

$$p_0(h) = \frac{\ell_0}{(1+\lambda)(1-\ell_0)} \tag{28}$$

then (3) has a unique root $z_h^* \in \overline{U}(L_h(x^*), r_0)$.

By (20), (21) and (23) we get

$$\ell_0 \leqq d^2 \ell \| F_h(L_h(x^*)) - \overline{L}_h(F(x^*)) \| \leqq d^2 \ell c_0 h^q < e$$

and

$$r_0 \leqq m d c_0 h^q \leqq w, \tag{29}$$

which shows that (27) and (28) hold for all h satisfying (23).

Thus (22) follows from

$$\| z_h^* - L_h(x^*) \| \leqq r_0 \leqq m d c_0 h^q. \tag{30}$$

By applying Theorem 1 to (3) we see that the Newton sequence (4) converges to z_h^* if

$$\| L_h(x_0) - z_h^* \| < \left(\frac{e}{\ell \| F_h'(z_h^*)^{-1} \|} \right)^{1/\lambda} \tag{31}$$

$$\overline{U}(z_h^*, \| L_h(x_0) - z_h^* \|) \subset \overline{U}(L_h(x^*), w). \tag{32}$$

But (32) holds if

$$\| z_h^* - L_h(x^*) \| + \| L_h(x_0) - z_h^* \| \leqq w, \tag{33}$$

and by (18) and (30) we have

$$\begin{aligned} \| L_h(x_0) - z_h^* \| &\leqq \| L_h(x_0) - L_h(x^*) \| + \| L_h(x^*) - z_h^* \| \leqq \\ &\leqq b \| x_0 - x^* \| + m d c_0 h^q. \end{aligned} \tag{34}$$

Hence (33) is satisfied if

$$b \| x_0 - x^* \| + 2 m d c_0 h^q \leqq w. \tag{35}$$

Since,

$$F_h'(z_h^*) = F_h'(L_h(z^*))[I - F_h'(L_h(z^*))^{-1}(F_h'(L_h(x^*)) - F_h'(z_h^*))]$$

using (17), (14) and (30) we get

$$\|F_h'(z_h^*)^{-1}\| \leqq \frac{\|F_h'(L_h(z^*))^{-1}\|}{1-\ell\|F_h(L_h(z^*))^{-1}\|\|L_h(x^*) - z_h^*\|^\lambda} \leqq \frac{d}{1-\ell d(mdc_0h^q)^\lambda}. \tag{36}$$

Thus, (31) holds when

$$b\|x_0 - x^*\| + 2mdc_0h^q < \left[\frac{e}{\ell}\left(\frac{1-\ell d(mdc_0h^q)^\lambda}{d}\right)\right]^{1/\lambda}. \tag{37}$$

By setting,

$$h_2 = \min\left[\left(\frac{w}{4mdc}\right)^{1/q}, \left(\frac{1}{4mdc}\right)^{1/q}\left(\frac{e(1-\ell dw^\lambda)}{\ell d}\right)^{1/\lambda q}\right], \tag{38}$$

and

$$r_2 = \min\left[\frac{w}{2b}, \frac{1}{2b}\left(\frac{e(1-\ell dw^\lambda)}{\ell d}\right)^{1/\lambda}\right], \tag{39}$$

it can easily be verified that (34) and (37) hold for all $h \in (0, h_2]$ and $x_0 \in$ $\in U(x^*, r_2)$. That is, for these h and x_0, the sequence (4) converges to z_h^*.

Let us now define the function v by

$$v = v(h) = c_2h^q, \quad c_2 > 0. \tag{40}$$

We now prove that for $h \in (0, h_1)$ and $x_0 \in \overline{U}(x^*, r_1)$ and all $n = 0, 1, \ldots$ the estimate

$$\|z_n^h - L_h(x_n)\| \leqq v \tag{41}$$

holds, where

$$h_1 = \min\left[h_0, h_2, \left(\frac{(C-B)^2}{4AD}\right)^{1/q}, \left(\frac{-T}{D}\right)^{1/2}\left(\frac{1}{\ell dc_2^\lambda}\right)^{1/q\lambda}\right] \tag{42}$$

and

$$r_1 = \min(r_2, r^*). \tag{43}$$

We use induction. For $n = 0$ (41) is trivially true.

Consider the identity,

$$z_{i+1}^h - L_h(x_{i+1}) = F_h'(z_i^h)^{-1}\Big\{[F_h'(z_i^h)(z_i^h - L_h(x_i)) - F_h(z_i^h) + F_h(L_h(x_i))] +$$

$$+[(F_h'(z_i^h) - F_h'(L_h(x_i))L_h(F'(x_i)^{-1}F(x_i))] + \tag{44}$$

$$+[F_h'(L_h(x_i)L_h(F'(x_i)^{-1}F(x_i)) - \overline{L}_h(F(x_i))] + [\overline{L}_h(F(x_i)) - F_h(L_h(x_i))]\Big\}.$$

As in (36) we can obtain

$$\|F_h'(z_i^h)^{-1}\| \leqq \frac{d}{1-\ell d v^\lambda}. \tag{45}$$

Using a standard argument we have that

$$\|F_h'(z_i^h)(z_i^h - L_h(x_i)) - F_h'(z_i^h) + F_h(L_h(x_i))\| \leqq$$
$$\leqq \frac{1}{\lambda+1}\ell\|z_i^h - L_h(x_i)\|^{\lambda+1} \leqq \frac{1}{\lambda+1}\ell v^{\lambda+1}. \tag{46}$$

Also,

$$\|(F_h'(z_i^h) - F_h'(L_h(x_i))(L_h(F'(x_i)^{-1}F(x_i))\| \leqq$$

$$\leqq \ell b\|z_i^h - L_h(x_i)\|^\lambda\|x_i - x_{i+1}\| \leqq 2\ell b v^\lambda\|x_0 - x^*\| \leqq 2\ell b v^\lambda r_1 \tag{47}$$

(since by Theorem 1 $\|x_{i+1} - x^*\| \leqq \|x_i - x^*\|$).

Finally, from (20) and (21) we obtain

$$\|F_h'(L_h(x_i))L_h(F'(x_i)^{-1}(F(x_i)) - \overline{L}_h F(x_i)\| \leqq c_1 h^q \leqq ch^q \tag{48}$$

and

$$\|\overline{L}_h F(x_i) - F_h(L_h(x_i))\| \leqq c_0 h^q \leqq ch^q. \tag{49}$$

Using the above estimates in (44) we obtain that

$$\|z_{i+1}^h - L_h(x_{i+1})\| \leqq \frac{d}{1-\ell d v^\lambda}\Big[\frac{1}{\lambda+1}\ell v^{\lambda+1} + 2\ell b v^\lambda r^* + 2ch^q\Big]. \tag{50}$$

Define the real functions f and g by

$$f(v) = Av^{\lambda+1} - Bv + Cv^\lambda + Dh^q \tag{51}$$

and

$$g(v) = Av^2 + (C - B)v + Dh^q. \tag{52}$$

By the choice of r_1 and h_1

$$C < B, \tag{53}$$

$$(C - B)^2 - 4ADh^q > 0, \tag{54}$$

and f has two positive solutions.

Therefore, the function g has a minimum at

$$v_m = \frac{B - C}{2A} \tag{55}$$

and

$$f(v_m) = T + Dh^q \tag{56}$$

which according to (24) and the choice of h is negative. Since $f(v)$ is continuous, $f(0) > 0$ and $f(v) > 0$ for v sufficiently large we are assured that $f(v)$ has two positive solutions. Denote by v_1 the smallest positive root. Then the right-hand side of (50) is equal to v_1.

Moreover,

$$v_1[Av_1^\lambda - B + Cv_1^{\lambda-1}] = -Dh^q \tag{57}$$

or

$$v_1 = \frac{D}{B - Av_1^\lambda - Cv_1^{\lambda-1}} h^q \tag{58}$$

with

$$B - Av_1^\lambda - Cv_1^{\lambda-1} > 0. \tag{59}$$

By (59), there exist v_2, v_3 sufficiently close to v_1 with $v_2 \leqq v_1 \leqq v_3$ such that

$$B - Av_3^\lambda - Cv_2^{\lambda-1} > 0. \tag{60}$$

Therefore, by (58), we obtain that

$$v_1 \leqq \frac{D}{B - Av_3^\lambda - Cv_2^{\lambda-1}} h^q = c_2 h^q \tag{61}$$

by setting

$$c_2 = \frac{D}{B - Av_3^{\lambda} - Cv_2^{\lambda-1}}.$$

This proves (25) since, we have

$$\|z_n^h - L_h(x_n)\| \leqq v_1 \leqq c_2 h^q. \tag{62}$$

Finally, by (30), (41), and (62), we get

$$\begin{aligned} \|(z_n^h - z_h^*) - L_h(x_n - x^*)\| &\leqq \|z_n^h - L_h(x_n)\| + \|z_h^* - L_h(x^*)\| \leqq \\ &\leqq mdc_0 h^q + c_2 h^q = c_3 h^q \end{aligned} \tag{63}$$

by setting $c_3 = mdc_0 + c_2$, which shows (26) and that completes the proof of the theorem.

We can now prove the following to justify the claims made in the introduction.

THEOREM 4. *Assume:*
(a) *the hypotheses of Theorem 3 are true;*
(b) *there exists a* $\delta > 0$ *such that*

$$\liminf_{h>0} \|L_h(u)\| \geqq \delta \|u\| \quad \text{for each} \quad u \in W^*. \tag{64}$$

Then for some $\overline{r} \in (0, r_1)$, *and for any fixed* $\varepsilon > 0$ *and* $x_0 \in U(x^*, \overline{r})$ *there exists a constant* $\overline{h}$ *depending on* ε *and* z_0 *with* $\overline{h} \in (0, h_1]$ *such that*

$$|\min\{n \geqq 0, \|x_n - x^*\| < \varepsilon\} - \min\{n \geqq 0, \|z_n^h - z_h^*\| < \varepsilon\}| \leqq 1 \tag{65}$$

for all $h \in (0, \overline{h}]$.

PROOF. Let k be the unique integer defined by

$$\|x_{k+1} - x^*\| < \varepsilon \leqq \|x_i - x^*\| \tag{66}$$

and $h_3 > 0$ such that

$$\|L_h(x_i - x^*)\| \geqq \delta \|x_i - x^*\|, \quad \text{with} \quad 0 < h < h_3. \tag{67}$$

Set

$$\overline{r} = \min\left(r_1, \frac{1}{2}\left(\frac{\overline{b}}{a + d\ell\overline{b}}\right)^{1/\lambda}\right), \tag{68}$$

$$a = \frac{d\ell}{1+\lambda}, \qquad \bar{b} = \min\left(b, \frac{1}{2b}, \frac{\delta}{2}\right), \tag{69}$$

and

$$\bar{h} = \min\left[h_1, h_3, \left(\frac{\bar{b}}{a + d\ell\bar{b}}\right)^{1/\lambda q}\left(\frac{1}{2mdc_0}\right)^{1/q}, \left(\frac{\bar{b}\varepsilon}{c_3}\right)^{1/q}\right]. \tag{70}$$

We will prove the theorem for the above choices of $\bar{r}$ and $\bar{h}$.

By (63) and (70) we obtain that

$$\|z_{i+1}^h - z_h^*\| \leqq \|L_h(x_{i+1} - x^*)\| + c_3h^q \leqq b\varepsilon + c_3h^q \leqq 2b\varepsilon, \tag{71}$$

and from (34), (68), (70) and (8)

$$\begin{aligned} \|z_{i+2}^h - z_h^*\| &\leqq \frac{d\ell\|z_{i+1} - z_h^*\|^{1+\lambda}}{(1+\lambda)[1 - d\ell\|z_{i+1} - z_h^*\|^{\lambda}]} \leqq \\ &\leqq \left(\frac{d\ell}{1+\lambda}\right)\frac{\|z_0^h - z_h^*\|^{\lambda}}{(1 - d\ell\|z_0^h - z_h^*\|^{\lambda})}\|z_{i+1} - z_h^*\| \leqq 2\bar{b}b\varepsilon < \varepsilon. \end{aligned} \tag{72}$$

By (67) and (63) we get

$$\varepsilon \leqq \|x_i - z^*\| \leqq \frac{1}{\delta}\|L_h(x_i - x^*)\| \leqq \frac{1}{\delta}(\|z_i^h - z_h^*\| + c_3h^q) \tag{73}$$

or

$$\|z_i^h - z_h^*\| \geqq \delta\varepsilon - c_3h^q \geqq \delta\varepsilon - \frac{\delta\varepsilon}{2} = \frac{\delta\varepsilon}{2}. \tag{74}$$

If $\|z_{i-1}^h - z_h^*\| < \varepsilon$, then as in (72) we get

$$\|z_i^h - z_h^*\| < \frac{\bar{b}\varepsilon}{2} \leqq \frac{\delta\varepsilon}{2} \tag{75}$$

which contradicts (75). That is,

$$\|z_{i-1}^h - z_h^*\| \geqq \varepsilon. \tag{76}$$

The result now follows from (66), (72), and (76).

REMARKS. (a) The condition (64) follows from

$$\lim_{h\to 0}\|L_h(u)\| = \|u\|, \quad u \in W^*. \tag{77}$$

(b) For some discretizations we have

$$\lim_{h\to 0}\|L_h(u)\| = \|u\| \quad \text{uniformly for } u \in W^*. \tag{78}$$

Both conditions above hold in many discretization studies [1]–[3], [5]–[7], [9], [12].

The following result is now immediate.

COROLLARY. *Assume:*
(a) *the hypotheses of Theorem 3 are satisfied;*
(b) *the condition* (78) *holds uniformly for* $u \in W^*$.

Then there exists $\overline{r}_1 \in (0, r_1]$ *and, for any fixed* $\varepsilon > 0$, *some* $\overline{h}_1 = \overline{h}_1(\varepsilon) \in \in (0, h_1]$ *such that*

$$|\min\{n \geqq 0, \|x_n - x^*\| < \varepsilon\} - \min\{n \geqq 0, \|z_n^h - z_h^*\| < \varepsilon\}| \leqq 1$$

holds for all $h \in (0, \overline{h}_1]$ *and all* $z_0 \in \overline{U}(x^*, \overline{r}_1)$.

EXAMPLE. Consider the differential equation

$$y'' + y^{1+\lambda} = 0, \quad \text{for} \quad \lambda \in (0,1)$$
$$y(0) = y(1) = 0.$$

Define the operator

$$F : M \subset C^2[0,1] \to C[0,1] \times \mathbf{R}^2,$$

$$F(y) = \{y'' + y^{1+\lambda};\ 0 \leqq x \leqq 1,\ y(0), y(1)\}.$$

Assume that M is such that the equation $F(y) = 0$ has a unique solution $x^* \in M$ and set

$$U(x^*, w) = \{(x_1^*, x_2^*, x_3^*) \in \mathbf{R}^3;\ 0 \leqq x_1^* \leqq 1, |x_2^* - x^*(x_1^*)| \leqq w, |x_3 - x^{*\prime}(x_1^*)| \leqq w\}.$$

It can easily be seen that $x^* \in C^3[0,1]$.

The Fréchet derivative of F is given by

$$F'(y)u = \{u'' + (1+\lambda)y(\overline{t}_n)^\lambda u, 0 \leqq x, \overline{t}_n \leqq 1, u(0), u(1)\}$$

and hence Newton's iteration becomes

$$x''_{n+1} = -x_n^{1+\lambda} + (1+\lambda)x_n^\lambda(\overline{t}_n)(x_n - x_{n+1})$$

with

$$x_{n+1}(0) = x_{n+1}(1) = 0.$$

Define the norm on $C^m[0,1]$, $m \geqq 0$ with

$$\|u\| = \{(\max|u^i(x)|, 0 \leqq x \leqq 1, i = 0, 1, \ldots, m\}.$$

Choose $x_0 \in C^2[0,1]$ then $x_{n+1} \in C^3[0,1]$, $n = 0, 1, 2, \ldots$. We will assume also that $x_0 \in C^3[0,1]$. By the convergence of x_n to x^* in the norm of $C^2[0,1]$, there exists $K > 0$ such that

$$x_n \in W_k = \{x \in C^3[0,1]; \sup_t |x^{(i)}(t)| \leqq K, i = 0, 1, 2, 3\},\ n = 0, 1, 2, \ldots.$$

By choosing sufficiently large K we assume

$$x^* \in W_K, \; x_n - x^* \in W_K \quad \text{and} \quad x_n - x_{n+1} \in W_K, \; n = 0, 1, \ldots .$$

We now divide the interval $[0, 1]$ into n subintervals and set $h = \frac{1}{n}$. We denote the points of subdivision by

$$p_0 = 0 < p_1 < \ldots < p_n = 1$$

with the corresponding values of the function $y_i = y(p_i)$, $i = 0, 1, 2, \ldots, n$.

A simple approximation for the derivative at these points is

$$y_i'' \simeq \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2}, \quad i = 1, 2, \ldots, n-1.$$

Since $y_0 = y_n = 0$ this leads to the following system of nonlinear equations

$$h^2 y_1^{1+\lambda} - 2y_1 + y_2 = 0,$$

$$y_{i-1} + h^2 y^{1+\lambda} - 2y_i + y_{i+1} = 0, \quad i = 2, 3, \ldots, n-1,$$

$$y_{n-2} + h^2 y_{n-1}^{1+\lambda} - 2y_{n-1} = 0.$$

We therefore have an operator $H : \mathbf{R}^{n-1} \to \mathbf{R}^{n-1}$ whose Fréchet-differential may be written as

$$H'(y) = \begin{bmatrix} (1+\lambda)h^2 y_1^{\lambda} - 2 & 1 & 0 & \ldots & 0 \\ 1 & (1+\lambda)h^2 y^{\lambda} - 2 & 1 & & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \ldots \quad 0 & 1 & & (1+\lambda)h^2 y_{n-1}^{\lambda} - 2 \end{bmatrix}.$$

Choose $\lambda = \frac{1}{2}$ for simplicity and let $x \in \mathbf{R}^{n-1}$ with norm given by

$$\|x\| = \max_{1 \leqq j \leqq n-1} |x_j|.$$

The corresponding norm on $Q \in \mathbf{R}^{n-1} \times \mathbf{R}^{n-1}$ is

$$\|Q\| = \max_{1 \leqq j \leqq n-1} \sum_{k=1}^{n-1} |Q_{jk}|.$$

Then for all $y, z \in \mathbf{R}$ with $|y_i| > 0$, $|z_i| > 0$, $i = 1, 2, \ldots, n-1$

$$\|H'(y) - H'(z)\| = \left\| \operatorname{diag}\left\{ \frac{3}{2} h^2 (y_j^{1/2} - z_j^{1/2}) \right\} \right\| = \frac{3}{2} h^2 \max_{1 \leqq j \leqq n-1} |y_j^{1/2} - z_j^{1/2}| \leqq$$

$$\leqq \frac{3}{2} h^2 [\max_{1 \leqq j \leqq n-1} |y_j - z_j|]^{1/2} = \frac{3}{2} h^2 \|y - z\|^{1/2}.$$

Here $\ell = \frac{3}{2}h^2$ and $\lambda = \frac{1}{2}$, therefore the results in [2], [3], [5]–[7], [10]–[12] cannot be applied here. As in [2] the discretization method $\{T_h, L_h, \overline{L}_n\}$ is defined as follows:

$$G_h = \{p_i = ih, i = 0, 1, \dots, n\},\ G_h^0 = G_h \setminus \{0, 1\},$$

$$E_h^1 = \{\eta\colon G_h \to \mathbf{R}\},\ \eta_i = \eta(p_i),\ i = 0, 1, \dots, n,$$

$$E_h = \{(\eta, a, b); \eta \in G_h^0 \to \mathbf{R}, a, b, \in \mathbf{R}\},$$

$$L_h(y) = y/G_h,\ \overline{L}_h(y, a, b) = (y/G_h^0, a, b),$$

$$F_h(\eta) = \left\{\frac{\eta_{i+1} - 2\eta_i + \eta_{i-1}}{h^2} + \eta_i^{3/2};\ i = 1, 2, \dots, n-1, \eta_0, \eta_n\right\}.$$

The following norms are used in the corresponding spaces

$$\|y\| = \max\{|y^i(x)|, 0 \leqq x \leqq 1, i = 0, 1, 2\}, \qquad y \in C^2[0, 1]$$

$$\|\gamma\| = \max\{|u(x)|, a, b; 0 \leqq x \leqq 1\}, \qquad \gamma = (u, a, b) \in C[0, 1] \times \mathbf{R}^2$$

$$\|\eta\| = \max\left\{|\eta_i|, \left|\frac{\eta_{i+1} - 2\eta_i + \eta_{i-1}}{h^2}\right|,\ i = 1, 2, \dots, n-1\right\}, \quad \eta \in E_h^1$$

$$\|\xi\| = \max\{|a|, |b|, |\eta_i|, i = 1, 2, \dots, n-1\}, \qquad \xi = (\eta, a, b) \in E_h.$$

It can now easily be seen that (18) is satisfied for $b = 1$ and (20), (21) are satisfied with $q = 3/2$.

Moreover, we can easily see with the above norms that

$$\|L_h(u)\| \leqq \|u\| \leqq \|L_h(u)\| + K\left(\frac{1}{6}h + 1\right)h \quad \text{for} \quad u \in W_K,$$

that is, (78) is satisfied.

Therefore, Theorem 3 and the Corollary may now apply.

Further examples on differential equations can be found in the references.

REFERENCES

[1] Allgover, E. L., McCormick, S. F. and Pryor, D. V., A general mesh independence principle for Newton's method applied to second order boundary value problems, *Computing* **23** (1979), 233–246. *MR* **83c**:65176

[2] Allgover, E. L., Böhmer, K., Potra, F.-A. and Rheinboldt, W. C., A mesh-independence principle for operator equations and their discretizations, *SIAM J. Numer. Anal.* **23** (1986), 160–169. *MR* **87h**:65107

[3] Balázs, M. and Goldner, G., On the method of the chord and on a modification of it for the solution of nonlinear operator equations, *Stud. Cerc. Mat.* **20** (1968), 981–990. Zbl **174**, 465 and *MR* **41** #6390

[4] Collatz, L., *The numerical treatment of differential equations*, 3rd edition, Die Grundlehren der mathematischen Wissenschaften, Bd. 60, Springer-Verlag, Berlin–Göttingen–Heidelberg, 1960. *MR* **22** #322

[5] JANKÓ, B., Sur la théorie unitaire des méthodes d'iteration pour la résolution des équations opérationnelles nonlinéaires. I, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **6** (1961), 301–311. *MR* **27** #1828

[6] JANKÓ, B., Sur la résolution des équations opérationnelles nonlinéaires, *Mathematica (Cluj)* **7** (**30**) (1965), 257–262. *MR* **34** #2146

[7] JANKÓ, B., On a unified theory of iteration methods for solving nonlinear operator equations. II, *Ann. Univ. Sci. Budapest. Eötvös Sect. Comput.* **6** (1985), 183–189. *MR* **89h**:47102

[8] KRASNOSELSKIĬ, M. A. et al., *Approximate solution of operator equations*, Wolters–Noordhoff Publishing, Groningen, 1972. *MR* **52** #6515

[9] ORTEGA, J. M. and RHEINBOLDT, W. C., *Iterative solutions of nonlinear equations in several variables*, Academic Press, New York–London, 1970. *MR* **42** #8686

[10] POTRA, F.-A. and RHEINBOLDT, W. C., On the mesh-independence principle for discretizations of nonlinear operator equations, Institute for Comp. Mat. and Appl., Techn. Report ICMA-84-74, Univ. Pittsburgh, Pittsburgh, PA, June 1984.

[11] RALL, L. B., A note on the convergence of Newton's method, *SIAM J. Numer. Anal.* **11** (1974), 34–36. *MR* **49** #8339

[12] RHEINBOLDT, W. C., An adaptive continuation process for solving systems of nonlinear equations, *Mathematical models and numerical methods* (Papers, Fifth Semester, Stefan Banach Internat. Math. Centre, Warsaw, 1975), Banach Center Publ., 3, PWN, Warsaw, 1978, 723–751. *MR* **83k**:65045

[13] VAINIKKO, G. M., Galerkin's perturbation method and the general theory of approximate methods for nonlinear equations, *USSR Comp. Math. and Math. Phys.* **4** (1967), 723–751. See also *Ž. Vyčisl. Mat. i Mat. Fiz.* **7** (1967), 723–751. *MR* **36** #1095

(Received May 14, 1990)

DEPARTMENT OF MATHEMATICS
NEW MEXICO STATE UNIVERSITY
LAS CRUCES, NM 88003
U.S.A.

Current address:

DEPARTMENT OF MATHEMATICS
CAMERON UNIVERSITY
LAWTON, OK 73505–6377
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 417–426

NEWTON-LIKE METHODS AND NONDISCRETE MATHEMATICAL INDUCTION

I. K. ARGYROS

Abstract

The method of nondiscrete mathematical induction is applied to Newton-like methods. The method yields a simple proof of the convergence and generally better error bounds than previously obtained.

1. Introduction

Consider the equation

$$F(x)=0, \quad x\in X, \quad 0\in Y \tag{1}$$

where F is a nonlinear operator mapping some subset D of a real Banach space X into a subset of a real Banach space Y. It is known that an iteration of the form

$$x_{n+1}=P(x_n) \tag{2}$$

with

$$P(x_n)=x_n-A_n^{-1}F(x_n) \tag{3}$$

is called a Newton-like method and the fixed points x^* of P are roots of equation (1). Here $\{A_n\}$ denotes a sequence of invertible linear operators. This is plainly too general and what is really implicit in the title is that A_n should be a conscious approximation to $F'(X_n)$ since when $A_n=F'(X_n)$, the method is the obvious generalization of the classical Newton–Kantorovich method.

Sufficient conditions for the convergence of the sequence $\{x_n\}$, $n=0,1,2,\ldots$ generated by (2) to a root x^* of (1) as well as estimates for the distances $\|x_n-x^*\|$, have been given by several authors.

Most of this work can be found in the excellent papers by W. Rheinboldt [10], [11], J. Dennis [2], and the references there.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 65H10.

Key words and phrases. Newton-like methods, Banach space, nondiscrete mathematical induction.

Akadémiai Kiadó, Budapest

Here motivated by the work of F. Potra and V. Ptak for Newton's method [6], [7], [8], we apply the method of nondiscrete mathematical induction to the Newton-like method (2).

The method yields a simple proof of the convergence and generally better estimates for the distances $\|x_n - x^*\|$ than the ones given in [2].

Our results reduce to the ones obtained in [6] for Newton's method when $A_n = F'(x_n)$, $n = 0, 1, 2, \ldots$ in (2).

To make the paper self-contained we state some of the results obtained in [2] and [6].

2. Basic results

We will need the following version of Theorem 2.6 in [2].

THEOREM 1. *Assume*

(1) *the Fréchet-derivative F' of F exists and $F' \in \mathrm{Lip}_k D_0$, where D_0 is the closure of an open convex set and $D_0 \subset D$;*

(2) *for every n with $x_0, x_1, \ldots, x_n \in D_0$ and given by*

$$x_{n+1} = x_n - A_n^{-1} F(x_n), \quad n = 0, 1, 2, \ldots$$

there is an invertible $A_n \in L(X, Y)$ and a positive real number a_n such that

$$\|A_n^{-1}\| \leqq a_n^{-1}; \tag{4}$$

(3) *for $\sigma \geqq 1$, $\Delta > 0$, both independent of n,*

$$\|F'(x_n) - A_n\| \leqq a_n + \sigma k \sum_{j=1}^{n} \|x_j - x_{j-1}\| - \Delta; \tag{5}$$

(4) *the sequence $\{a_n\}$, $n = 0, 1, 2, \ldots$ is uniformly bounded above;*

(5) *the following estimates are true:*

$$\frac{1}{2} \geqq h \approx \frac{\sigma k \|A_0^{-1} F(x_0)\| a_0}{\Delta^2}, \tag{6}$$

$$\cup(x_0, r_0) \subset D_0, \quad \text{where} \quad r_0 = \frac{1 - \sqrt{1 - 2h}}{\sigma k} \Delta. \tag{7}$$

Then the sequence $\{x_n\}$, $n = 0, 1, 2, \ldots$ given by (2) is well defined and converges to a root x^* of the equation

$$F(x) = 0.$$

The error satisfies the estimate

$$\|X_{n+1} - x^*\| \leqq r_0 - t_n - a_n^{-1}\left(\frac{1}{2}\sigma k t_n^2 - \Delta t_n\right) + a_0\|A_0^{-1}F(x_0)\|; \tag{8}$$

where

$$t_0 = 0 \text{ and } t_{n+1} = (2a_0)^{-1}kt_n^2 + \delta t_n + \|A_0^{-1}F(x_0)\|, \quad n = 1, 2, \ldots \tag{9}$$

with

$$\delta \equiv \|F'(x_0) - A\| \Big/ a_0. \tag{10}$$

Moreover assume:

(6) *The following estimates are true:*

$$\delta < 1, \tag{11}$$

$$\frac{1}{2} > h' \equiv \frac{k\|A_0^{-1}F(x_0)\|}{a_0(1-\delta)^2}, \tag{12}$$

$$\cup(x_0, r_0') \subset D_0, \quad \text{where} \quad r_0' \equiv \frac{1 - \sqrt{1-2h'}}{k}(1-\delta)a_0. \tag{13}$$

Then x^* is the only root of F in $\cup(x_0, r_1') \cap D_0$.

We can now prove the theorem:

THEOREM 2. *Assume*

(1) *the Fréchet-derivative F' of F exists and $F' \in \mathrm{Lip}_k D_0$, where D_0 is the closure of an open convex set and $D_0 \subset D$;*

(2) *the linear operator $A_0 \in L(X, Y)$ is invertible with*

$$\|A_0^{-1}F(x_0)\| \leqq \eta, \tag{14}$$

$$\|A_0^{-1}\| < \beta \equiv a_0^{-1} \tag{15}$$

for some $\eta, \beta > 0$;

(3) *there exist real nonnegative sequences $\{q_n\}$ and $\{p_n\}$, $n = 0, 1, 2, \ldots$ such that for every n for which $x_0, x_1, x_2, \ldots, x_n$ as defined by (2) are in D_0,*

$$\|F'(x_n) - A_n\| \leqq p_n + q_n\left(\sum_{j=1}^{n} \|x_j - x_{j-1}\|\right) \equiv B_n; \tag{16}$$

$$q_n \leqq q \tag{17}$$

$$p_n \leqq p; \tag{18}$$

(4) *the following estimates hold*

$$\beta p_0 + 2\beta p < 1; \tag{19}$$

$$\frac{1}{2} \geqq h \equiv \frac{(2q+k)\beta\eta}{(1-2\beta p-\beta p_0)^2}; \tag{20}$$

and

$$\cup(x_0, r_0) \subset D_0, \quad \text{where} \quad r_0 = \frac{1-\sqrt{1-2h}}{\beta(2q+k)}(1-2\beta p-\beta p_0). \tag{21}$$

Then the sequence $\{x_n\}$ *given by* (2) *is well defined in* $\cup(x_0, r_0)$ *and converges to a unique root of* F *in* $\overline{\cup}(x_0, R_1)$, *where*

$$R_1 = \frac{1-\sqrt{1-2h'}}{\beta k}(1-\beta p_0) \tag{22}$$

with

$$h' = \frac{\beta k\eta}{(1-\beta p)^2}. \tag{23}$$

Moreover if

$$h' < \frac{1}{2} \tag{24}$$

then the root x^* *of* F *is unique in* $D_0 \cap \cup(x_0, R_2)$ *where*

$$R_2 = \frac{1+\sqrt{1-2h'}}{\beta k}(1-\beta p_0). \tag{25}$$

Finally, the following estimate is also true:

$$\|x_{n+1} - x^*\| \leqq \tag{26}$$
$$\leqq r_0 - t_n - \left(\frac{1}{2}\beta(k+2q)t_n^2 - (1-\beta(p_0+2p)t_n+\eta)/(1-\beta(p_n+p_0)-\beta(k+q)t_n\right)$$

with $t_0 = 0$, $t_1 = \eta$ *and* t_n, $n = 2, 3, \ldots$, *given by* (9).

PROOF. We will make use of Theorem 1. Assume that

$$x_0, x_1, \ldots, x_n \in \cup(x_0, r_0) \quad \text{and} \quad \sum_{j=1}^{n} \|x_j - x_{j-1}\| < r_0.$$

Then

$$\|A_n - A_0\| \leqq \|A_n - F'(x_n)\| + \|F'(x_n) - F'(x_0)\| + \|F'(x_0) - A_0\| \leqq$$

$$\leqq p_n + q_n\Big(\sum_{j=1}^{n} \|x_j - x_{j-1}\|\Big) + k\|x_n - x_0\| + p_0 \leqq$$

(27)

$$\leqq p_0 + p_n + (k+q)\sum_{j=1}^{n} \|x_j - x_{j-1}\| < p_0 + p + (k+q)r_0 \leqq$$

$$\leqq p_0 + p + (1 - 1 - 2h)(1 - 2\beta p - \beta p_0)/\beta \leqq \frac{1}{\beta}.$$

Hence, $\|A_0^{-1}A_n - I\| \leqq \beta(p_0+p_n)+\beta(k+q)\sum_{j=1}^{n} \|x_j - x_{j-1}\|$ and by the Banach lemma on invertible operators, A_n^{-1} exists and

(28)

$$\|A_n^{-1}\| \leqq a_n^{-1} \equiv \beta\Big(1 - \beta p_n - \beta p_0 - \beta(k+q)\sum_{j=1}^{n} \|x_j - x_{j-1}\|\Big)^{-1} \leqq a_0^{-1},\ n > 0.$$

We now need to find σ and Δ such that

$$q_n\Big(\sum_{j=1}^{n} \|x_j - x_{j-1}\|\Big) + p_n \leqq q\Big(\sum_{j=1}^{n} \|x_j - x_{j-1}\|\Big) + p_n \leqq$$

$$\leqq \Big(1 - \beta p_n - \beta p_0 - \beta(k+q)\sum_{j=1}^{n} \|x_j - x_{j-1}\|\Big)\beta^{-1} + \sigma k\sum_{j=1}^{n} \|x_j - x_{j-1}\| - \Delta.$$

It can easily be checked that for

(29)

$$\Delta \equiv \frac{1 - 2\beta p - \beta p_0}{\beta}$$

and

(30)

$$\sigma \equiv \frac{k+2q}{k}$$

the above inequality is satisfied.

The conclusions of the Theorem follow immediately now from Theorem 1.

DEFINITION. A function $\omega: T \equiv \{r \in \mathbf{R}, 0 < r < s, s \text{ fixed}\} \to T$ is called a rate of convergence on T if the series

(31)

$$\sigma(r) = \sum_{n=0}^{\infty} \omega^{(n)}(r)$$

is convergent for each $r \in T$, where the iterates $\omega^{(n)}$ of ω are defined as follows

$$\omega^{(0)}(r) = r, \quad \omega^{(n+1)}(r) = \omega(\omega^{(n)}(r)), \quad n = 0, 1, 2, \ldots . \tag{32}$$

From the definition of ω and σ it follows immediately that

$$\sigma(\omega(r)) = \sigma(r) - r, \quad r \in T. \tag{33}$$

EXAMPLE 1. By Lemma 2.2 in [6], it follows that the function

$$\omega(r) = \frac{r^2}{2(r^2 + a^2)^{1/2}}, \quad a \geqq 0 \tag{34}$$

is a rate of convergence on T and the corresponding function is given by

$$\sigma(r) = r - a + (r^2 + a^2)^{1/2}. \tag{35}$$

Moreover, the following estimates are true:

$$\omega^{(n)}(r) = \frac{2a[\theta(r)]^{2^n}}{1 - [\theta(r)]^{2^{n+1}}} \tag{36}$$

and

$$\sigma(\omega^{(n)}(r)) = \frac{2a[\theta(r)]^{2^n}}{1 - [\theta(r)]^{2^n}}, \quad n = 0, 1, 2, \ldots , \tag{37}$$

where

$$\theta(r) = \frac{(r^2 + a^2)^{1/2} - a}{r} < 1, \quad \text{for} \quad r > 0. \tag{38}$$

We will need the Theorem in [6].

THEOREM 3. *If we can attach to the pair (P, x_0) a rate of convergence ω on an interval T and a family of sets $Z(r) \subset X$, $r \in T$ such that the conditions*

$$x_0 \in Z(r_0) \quad \textit{for a certain} \quad r_0 \in T, \tag{39}$$

$$(r \in T \textit{ and } x \in Z(r)) \to P(x) \in U(x, r) \cap Z(\omega(r)) \tag{40}$$

are satisfied then the iteration

$$x_{n+1} = P(x_n), \quad n = 0, 1, 2, \ldots$$

is well defined and converges to a fixed point x^* of P such that

$$x_n \in Z(\omega^{(n)}(r_0)) \tag{41}$$

$$\|x_n - x_{n-1}\| \leqq \omega^{(n)}(r_0) \tag{42}$$

$$\|x_n - x^*\| \leqq \sigma(\omega^{(n)}(r_0)). \tag{43}$$

Moreover if for certain $n \in \{1, 2, \dots\}$

$$x_{n-1} \in Z(\|x_n - x_{n-1}\|) \tag{44}$$

then

$$\|x_n - x^*\| \leqq Q(\|x_n - x_{n-1}\|) \tag{45}$$

where we have denoted

$$Q(r) = \sigma(r) - r. \tag{46}$$

We can now prove the main result.

THEOREM 4. *Assume*
(1) *the hypotheses of Theorem 2 are satisfied;*
(2) *there exists a positive increasing continuous function* V *such that*

$$V(R_1) \geqq a_0, \tag{47}$$

$$V(\omega^{(n-1)}(r)) - [(B_{n-1} + B_n) + k\omega^{(n-1)}(r)] \geqq V(\omega^{(n)}(r)), \tag{48}$$

$$[V(\omega^{(n-1)}(r))]^{-1}\left[\frac{1}{2}k(\omega^{(n-1)}(r))^2 + B_n(\omega^{(n-1)}(r))\right] \leqq \omega^{(n)}(r) \tag{49}$$

for every $n = 1, 2, \dots$ *and* $r \geqq R_1$. *Then iteration* (2) *is well defined and converges to a unique solution* x^* *of the equation*

$$F(x) = 0$$

in $\cup(x_0, R_1) \cap D_0$.

The following relations are true:

$$x_n \in Z(\omega^{(n)}(R_1)) \tag{50}$$

$$\|x_n - x_{n-1}\| \leqq \omega^{(n)}(R_1) \tag{51}$$

$$\|x_n - x^*\| \leqq \sigma(\omega^{(n)}(R_1)) \tag{52}$$

$$\|x_n - x^a\| \leqq Q(\|x_n - x_{n-1}\|) \tag{53}$$

where ω, σ, Q *are as defined by* (34), (35) *and* (46), *respectively.*

Here,

$$a = \frac{\sqrt{1 - 2h'}}{\beta k}(1 - \beta p_0), \tag{54}$$

(55) $$\theta = \theta(R_1)$$

and

$$Z(\omega^{(n)}(r)) = \{x_n \in X \mid \|x_n - x_0\| \leqq \sigma(R_1) - \sigma(\omega^{(n)}(r))\},$$

A_n is boundedly invertible and

(56) $$\|A_n^{-1}\|^{-1} \geqq V(\omega^{(n)}(r)), \qquad \|A_n^{-1}F(x_n)\| \leqq \omega^{(n)}(r).$$

PROOF. According to Theorem 3, we need to show that conditions (39), (40) and (44) are satisfied. The hypotheses of the Theorem imply that $Z(R_1) = \{x_0\}$, so that (39) is satisfied. We will use induction to show that $X_{n+1} \in Z(\omega^{n+1}(r))$ if $x_j \in Z(\omega^{(j)}(r))$, $j = 0, 1, 2, \ldots, n$.

Since

$$x_{n+1} = x_n - A_n^{-1}F(x_n)$$

we have

$$\|x_{n+1} - x_0\| \leqq \|x_{n+1} - x_n\| + \|x_n - x_0\| \leqq$$

$$\leqq \omega^{(n)}(r) + \sigma(R_1) - \sigma(\omega^{(n)}(r)) = \sigma(R_1) - \sigma(\omega^{(n+1)}(r)).$$

$$\|A_{n+1}^{-1}\|^{-1} \geqq \|A_n^{-1}\|^{-1} - \|A_{n+1} - A_n\| \geqq$$

$$\geqq V(\omega^{(n)}(r)) - [\|A_n - F'(x_n)\| + \|F'(x_n) - F'(x_{n+1})\| + \|F'(x_{n+1} - A_{n+1}\|] \geqq$$

$$\geqq V(\omega^{(n)}(r)) - [B_n + B_{n+1} + k(\omega^{(n)}(r))] \geqq V(\omega^{(n+1)}(r))$$

(by (48)).

From the identity

(57)
$$F(x_{n+1}) = F(x_{n+1}) - F(x_n) - F'(x_n)(x_{n+1} - x_n) + (F'(x_n) - A_n)(x_{n+1} - x_n)$$

we get

$$\|A_{n+1}^{-1}F(x_{n+1})\| \leqq [V(\omega^{(n)}(r))]^{-1}\left[\frac{1}{2}k(\omega^{(n)}(r))^2 + B_{n+1}(\omega^{(n)}(r))\right] \leqq \omega^{(n+1)}(r)$$

(by (49)).

Thus conditions (39) and (40) are then satisfied.

Also, since $x_{n+1} \in Z(\omega^{(n-1)}(R_1))$ and the monotonicity of the functions σ and V we get

$$\|x_n - x_{n-1}\| = \|A_{n-1}^{-1}F(x_{n-1})\| \leqq \omega^{(n-1)}(R_1),$$

$$\|x_{n-1} - x_0\| \leqq \sigma(R_1) - \sigma(\omega^{(n-1)}(R_1)) \leqq \sigma(R_1) - \sigma(\|x_n - x_{n-1}\|),$$

$$\|A_{n-1}^{-1}\|^{-1} \geqq V(\omega^{(n-1)}(R_1)) \geqq V(\|x_n - x_{n-1}\|).$$

That is, (44) is also verified.

Finally, the proof of the Theorem can be completed if we use the continuity of F in (57) to obtain

$$F(x^*) = 0.$$

EXAMPLE 2. (a) The function V cannot easily be found at this generality. The conditions (47), (48), (49) may not even be satisfied for $r = R_1$. However, if V can be found then the relations (52), (53) indicate the possibility of obtaining sharper error bounds than previously known (see, e.g. [2], [10]).

(b) In the special case when $A_n = F'(x_n)$ then

$$p_n = p = q_n = q = 0 \quad \text{in (16)}, \quad n = 0, 1, 2, \ldots$$

and Theorem 2 reduces to the well-known Newton–Kantorovich theorem [3].

Moreover, it can easily be seen then as in [6] that the function

$$V(r) = a_0 - k(\sigma(R_1) - \sigma(r)) = k(r + (r^2 + a^2)^{1/2}) \tag{58}$$

satisfies conditions (47), (48), and (49).

(c) If, say, inequality (48) (or (49)) holds for every $n = 1, 2, \ldots$ with $r > 0$ fixed, then

$$\lim_{n \to \infty} B_n = 0$$

and by (16)

$$\lim_{n \to \infty} (F'(x_n) - A_n) = 0.$$

The above observation can lead to a remark similar to the one in (b).

However, in practice since the estimates given by (51), (52) and (53) will be computed until a finite $n = n_0$ and since the existence of x^* is guaranteed by Theorem 2, we will only require (48) and (49) to hold until η_0.

Another problem left then is the choice of V. A good candidate for V may then be the function given by (58).

REFERENCES

[1] ARGYROS, I. K., On Ptàk's analysis of Newton's method, *Bull. Austral. Math. Soc.* **39** (1988), 104–115.

[2] DENNIS, J. E., Toward a unified convergence theory for Newton-like methods, *Nonlinear Functional Anal. and Appl.*, (Proc. Advanced Sem., Math. Res. Center, Univ. of Wisconsin, Madison, Wis., 1970), Academic Press, 1971, 425–472. *MR* **43** #4286

[3] KANTOROVIČ, L. V. and AKILOV, G. P., *Functional analysis*, 2nd edition, Nauka, Moscow, 1977 (in Russian). *MR* **58** #23465

[4] MOORE, R. H., Newton's method and variations, *Nonlinear Integral Equations*, (Proc. Advanced Seminar conducted by Math. Research Center, U.S. Army, Univ. Wisconsin, Madison, Wis., 1963), ed. by P. Anselone, University of Wisconsin Press, Madison, Wis., 1964, 65–98. *MR* **29** #749

[5] ORTEGA, J. M. and RHEINBOLDT, W. C., *Iterative solution of nonlinear equations in several variables*, Academic Press, New York, 1970. *MR* **42** #8686

[6] POTRA, F.-A. and PTÁK, V., Sharp error bounds for Newton's process, *Numer. Math.* **34** (1980), 63–72. *MR* **81c**:65027

[7] POTRA, F.-A. and PTÁK, V., *Nondiscrete induction and iterative processes*, Research notes in mathematics, Vol. 103, Pitman, Boston, Mass.-London, 1984. *MR* **86i**:65003

[8] PTÁK, V., The rate of convergence of Newton's process, *Numer. Math.* **25** (1976), 279–285. *MR* **57** #18064

[9] PTÁK, V., What should be a rate of convergence? *RAIRO Anal. Numér.* **11** (1977), 279–286. *MR* **57** #14432

[10] RHEINBOLDT, W. C., A unified convergence theory for a class of iterative processes, *SIAM J. Numer. Anal.* **5** (1968), 42–63. *MR* **37** #1061

[11] RHEINBOLDT, W. C., *Numerical analysis of parametrized nonlinear equations*, University of Arkansas Lecture Notes in the Mathematical Sciences, Vol. 7, Wiley, New York, 1986. *MR* **87b**:65079

(Received May 14, 1990)

DEPARTMENT OF MATHEMATICS
NEW MEXICO STATE UNIVERSITY
LAS CRUCES, NM 88003
U.S.A.

Current address:

DEPARTMENT OF MATHEMATICS
CAMERON UNIVERSITY
LAWTON, OK 73505-6377
U.S.A.

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 427–429

NOTE ON ADDITIVE FUNCTIONS SATISFYING SOME CONGRUENCE PROPERTY. II

PHAM VAN CHUNG

Let $\mathcal{A}$ and $\mathcal{A}^*$ denote the set of integer-valued additive and completely additive functions, respectively. We shall denote by $\mathbf{N}$ resp. $\mathbf{Z}$ the set of positive integers and integers.

K. Kovács [2] proved that if $f \in \mathcal{A}^*$ and for some $a > 0$, $b, c \in \mathbf{Z}$

$$f(an+b) \equiv c \,(\mathrm{mod}\, n)$$

then $f(n) = 0$ for all $(n, a) = 1$.

In [1] we get the same result for $f \in \mathcal{A}$ if $a = 1$. Here we prove the following generalization of the above result:

THEOREM. *Let $A > 0$, B and C be integers. If $f \in \mathcal{A}$ satisfies the condition*

$$f(An+B) \equiv C \,(\mathrm{mod}\, n) \quad \textit{for all} \quad n > \max\{0, -B/A\}, \tag{1}$$

then $f(n) = 0$ for all $n \in \mathbf{N}$ which are coprime to A.

PROOF. We shall prove the theorem in three cases according to $B > 0$, $B = 0$ and $B < 0$.

Case I. If $B > 0$, then replacing n by $B^2 n$ in (1), we have

$$f(AB^2 n + B) \equiv C \,(\mathrm{mod}\, n) \tag{2}$$

which implies

$$f(ABn+1) \equiv C - f(B) \,(\mathrm{mod}\, n). \tag{3}$$

Using the method of our paper [1], one can deduce from (3) and (1) that

$$C = f(B), \tag{4}$$

$$f(ABn+1) \equiv 0 \,(\mathrm{mod}\, n) \tag{5}$$

1991 *Mathematics Subject Classification.* Primary 11A25.
Key words and phrases. Characterization of additive functions.

This paper is granted to a one-year scholarship (at the Eötvös University, Department of Algebra and Number Theory) for Viet-Nameses educated in Hungary.

and

$$f(n) = 0 \quad \text{for all} \quad (n, AB) = 1. \tag{6}$$

It remains to consider the case with $(n, A) = 1$ but $(n, B) \neq 1$. We may assume that $n = p^k$ where p is a prime and $k \in \mathbf{N}$.

Let us denote $p^\alpha \| B$ if $p^\alpha | B$ but $p^{\alpha+1} \nmid B$. We show first $f(p^k) = 0$ for all $0 \leqq k \leqq \alpha$. Since $(p, A) = 1$, there are infinitely many positive u for which

$$(Au + p^{\alpha-k}, AB) = 1. \tag{7}$$

By (1), (4) and (7), we get

$$f(B) \equiv f\Big(A\frac{B}{p^{\alpha-k}}u + B\Big) = f\Big(\frac{B}{p^{\alpha-k}}\Big) + f(Au + p^{\alpha-k}) \,(\mathrm{mod}\, u).$$

But (6) and (7) yield $f(Au + p^{\alpha-k}) = 0$, i.e. $f(B) \equiv f\Big(\frac{B}{p^{\alpha-k}}\Big) \,(\mathrm{mod}\, u)$. Thus we have

$$f(B) = f\Big(\frac{B}{p^{\alpha-k}}\Big) \quad \text{for all} \quad 0 \leqq k \leqq \alpha.$$

Since

$$f\Big(\frac{B}{p^{\alpha-k}}\Big) = f\Big(\frac{B}{p^\alpha}\Big) + f(p^k)$$

for all $0 \leqq k \leqq \alpha$, we obtain $f(p^\alpha) = f(p^k)$ for all $0 \leqq k \leqq \alpha$. The choice $k = 0$ implies

$$f(p^\alpha) = 0 \tag{8}$$

and so we have

$$f(p^k) = 0 \quad \text{for all} \quad 0 \leqq k \leqq \alpha. \tag{9}$$

Let us consider the case $k > \alpha$. To prove $f(p^k) = 0$ we show that

$$f(Bp^s) = f(B) \quad \text{for all} \quad s \in \mathbf{N}.$$

By $(p, A) = 1$, there exists a positive integer $D > 1$ such that $(D, AB) = 1$ and

$$p^s D = 1 + AT. \tag{10}$$

So by the theorem of Euler we get

$$D^{\varphi(A)m} \equiv 1 \,(\mathrm{mod}\, A). \tag{11}$$

(10) and (11) yield

$$p^s B D^{1+\varphi(A)m} \equiv B \,(\mathrm{mod}\, A). \tag{12}$$

From $f(D^{1+\varphi(A)m}) = 0$, the congruences (12) and (1) imply

$$f(p^s B) = f(p^s B) + f(D^{1+\varphi(A)m}) = f(B + ABI_m) \equiv f(B) \pmod{I_m}.$$

If $I_m \to \infty$, then we have

$$f(p^s B) = f(B),$$

which yields

$$f(p^{s+\alpha}) = f(p^\alpha) \quad \text{for all} \quad s \in \mathbf{N},$$

hence $f(p^k) = f(p^\alpha) = 0$ for all $k > \alpha$.

Case II. We can prove similarly as in [1].

Case III. If $B < 0$, then similarly to the Case III in [1] we get $f(n) = 0$ for all n coprime to A. If AB is odd, then we obtain $f(n) = 0$ for all n coprime to $2A$ only, but an analogous proof to the Case I with $p^s = 2^k$ implies also $f(2^k) = 0$.

REFERENCES

[1] Chung, Pham Van, Note on additive functions satisfying some congruence property. I, *Studia Sci. Math. Hungar.* **28** (1993), 359–362.

[2] Kovács, K., On additive functions satisfying some congruence properties, *Period. Math. Hungar.* **23** (1991), 227–231.

[3] Joó, I., On arithmetical functions satisfying some congruence properties, *Acta Math. Hungar.* (to appear).

(Received July 2, 1990)

KHOA TOAN
ĐAI HOC SU PHAM I
HANOI
VIETNAM

Present address:

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
ALGEBRA ÉS SZÁMELMÉLET TANSZÉK
MÚZEUM KRT. 6–8
H–1088 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 431–445

ÜBER KUGELSYSTEME UNTER GERÄUMIGKEITSBEDINGUNGEN

KATALIN BOGNÁR MÁTHÉ

Das Problem der dichtesten Packung kongruenter Kugeln ist bekanntlich für den dreidimensionalen euklidischen Raum E^3 noch ungelöst. Eine der besten oberen Abschätzungen für die Dichte d ein solcher Packung stammt von Rogers (1958). Er bewies die Vermutung $d \leqq d_n$, wonach die Dichte d einer jeden Packung von Kugeln mit Radius r in E^n nicht größer sein kann, als die Dichte d_n in einem regulären Simplex mit der Kantenlänge $2r$, das durch die Mittelpunkte von $n+1$ einander gegenseitig berührenden Kugeln vom Radius r bestimmt ist. In dem dreidimensionalen euklidischen Raum gilt nach Rogers $d \leqq 0,7796\ldots$. Diese Abschätzung für die Dichte d wurde von K. Böröczky auf $d \leqq 0,7784\ldots$ verbessert.[1] Es wird vermutet, daß die dichteste Kugelpackung die Dichte $d \leqq \frac{\pi}{\sqrt{18}} = 0,7404\ldots$ besitzt.

Bei den Untersuchungen von Kugelsystemen werden zumeist zusätzliche Eigenschaften gefordert; so werden Kugelpackungen mit Gitterförmigkeit, Schnurförmigkeit oder mit sonstigen Nebenbedingungen untersucht.

Die *Schnurförmigkeit* schwächt die Bedingung der Gitterförmigkeit ab, aber fordert immerhin eine gewisse Regularität. Das Problem wurde von L. Fejes Tóth und K. Böröczky aufgeworfen. Bei einer schnurförmigen Kugelpackung sind die Kugeln auf Geraden aufgefädelt. Die Abstände der Mittelpunkte in einer Reihe von Kugeln sind gleich, aber die Kugeln berühren einander nicht unbedingt. Nach einem Ergebnis von E. Makai kann die Dichte der dichtesten schnurförmigen Kugelpackung nicht größer sein, als die Dichte der dichtesten gitterförmigen Kugelpackung.

Eine große Anzahl von Kugelpackungen ist homogen oder zeigt in irgendeinem Sinne Regelmäßigkeiten. Ein solches Kugelsystem ist z.B. die würfelgitterförmige Kugelpackung. Diese Kugelsysteme können durch gewisse Nebenbedingungen als die Dichtesten charakterisiert werden. Eine solche Nebenbedingung ist z.B. *die Geräumigkeit.*

In der Ebene konstanter Krümmung hat J. Molnár den Begriff des geräumigen Kreissystems eingeführt [6], [7]. Der Name der „Geräumigkeit" stammt von L. Fejes Tóth.

1991 *Mathematics Subject Classification.* Primary 52C17.
Key words and phrases. Packing of spheres, density.

[1] Ein Vortrag in Seminar von Math. Inst. der Ungarischen Akademie der Wissenschaften (1980).

Akadémiai Kiadó, Budapest

Wir wählen einen Kreis K_i mit Mittelpunkt O_i des Kreissystems $\{K_i\}$, und betrachten die zu O_i nächstliegende Ecke E_i der Dirichlet-Voronoischen Zelle DV_i. Wir nennen den Abstand O_iE_i die Geräumigkeit des Kreises K_i. Man nennt ein Kreissystem $\{K_i\}$ geräumig, wenn die Abstände der DV_i Zellen-Ecken jedes Kreises K_i vom Mittelpunkt O_i in einen gewissen Sinne „groß" ist. Präziser gefaßt versteht man unter der Geräumigkeit eines Kreissystems $\{K_i\}$ das Infimum der Abstände O_iE_i bezüglich aller Kreise K_i.

Unter der Geräumigkeitsbedingung τ eines Kreissystems versteht man, daß die Geräumigkeit von $\{K_i\}$ mindestens τ ist.

Bei den Packungen der kongruenten Kreise mit Geräumigkeitsbedingungen gibt das eingeschriebene Kreissystem aller regulären Mosaike eine dichteste Kreispackung. Darüber hinaus gab J. Molnár [7] in der euklidischen Ebene für alle Geräumigkeitsbedingungen $\tau \geqq \frac{2}{\sqrt{3}}r$ die dichteste Packung von Kreisen mit Radius r an.

In den drei- und höher-dimensionalen Räumen konstanter Krümmung beschäftigte sich K. Böröczky mit geräumigen Kugelsystemen. Bei den Untersuchungen von Kugelsystemen kann man verschiedene Begriffe der Geräumigkeit einführen.

Wir betrachten ein Kugelsystem $\{K_i\}$ der Kugeln K_i vom Radius r im dreidimensionalen euklidischen Raum, und bezeichnen wir die Dirichlet–Voronoische Zelle der Kugel K_i mit Mittelpunkt O_i mit DV_i. Wir führen drei verschiedene Begriffe der Geräumigkeit ein.

Unter der *Eckengeräumigkeit* irgendeiner Kugel K_i von $\{K_i\}$ versteht man den Abstand der zu O_i nächstliegenden DV_i-Ecke vom Mittelpunkt O_i. Analog definiert man die *Kanten-*, bzw. *Flächengeräumigkeit* von K_i als den Abstand der zu O_i nächstliegenden DV_i-Kante, bzw. DV_i-Fläche vom Mittelpunkt O_i. Unter der entsprechenden Geräumigkeit eines Kugelsystems $\{K_i\}$ versteht man dann das Infinum der jeweiligen Geräumigkeiten von allen Kugeln K_i.

Im Falle einer Packung kongruenter Kugeln von Radius r sind die Ecken-, Kanten-, bzw. die Flächengeräumigkeiten bekanntlich mindestens $r\sqrt{\frac{3}{2}}$; $r\frac{2}{\sqrt{3}}$; bzw. r. Diese letzteren werden triviale Geräumigkeiten genannt.

Wenn eine Kugelpackung z.B. mit Geräumigkeitsbedingung r_3 angegeben wird, verlangt man, daß die Eckengeräumigkeit mindestens gleich r_3 ist, d.h. daß die Abstände der DV_i-Ecken vom Mittelpunkt O_i mindestens gleich r_3 sind. Analog kann man über eine Kantengeräumigkeitsbedingung r_2, bzw. über eine Flächengeräumigkeitsbedingung r_1 sprechen.

Nach Ergebnissen von K. Böröczky [1] vertreten alle regulären Mosaike im dreidimensionalen sphärischen, euklidischen und hyperbolischen Raum, deren Zellen regulären Polyeder sind, bei entsprechenden Geräumigkeitsbedingungen extreme Werte.

Die Eckengeräumigkeitsbedingung r_3 ist damit äquivalent, daß die Stützkugeln des Mittelpunktsystems der Kugelpackung mindestens den Radius r_3

haben.

Wenn bei einer Packung der kongruenten Kreise bzw. Kugeln nur die Eckengeräumigkeitsbedingung r_3 vorgeschrieben wird, dann nennt man ein solches Kreis- bzw. Kugelsystem auch ein ρ-System, wo $r = \rho + r$ ist. Die Untersuchung der ρ-Systeme hat J. Molnár [8] eingeführt.

M. Hollai [5] hat über die Dichte des gitterförmigen ρ-Systems der Kugeln genaue Schranken angegeben. M. Hollai hat zu einem jeden Wert

$$\rho \geqq r\left(\frac{\sqrt{3}}{2} - 1\right)$$

eine dichteste gitterförmige Packung angegeben.

Über ein quasi-geräumiges Kugelsystem im euklidischen Raum

Die Geräumigkeitsbedingungen können abgeschwächt werden. Man kann z.B. die Geräumigkeitsbedingung etwa nur für bestimmte ausgezeichnete Ecken der DV-Zellen einer Kugelpackung fordern. Diese speziellen Geräumigkeitsbedingungen nennen wir Quasi-Geräumigkeitsbedingung.

Werden wir uns einer Quasi-Geräumigkeitsbedingung zu! Zwei Ecken einer DV-Zelle heißen benachbart, wenn sie eine Kante beranden. Eine Kugelpackung erfüllt genau dann eine Quasi-Geräumigkeitsbedingung r_3^*, wenn von je zwei benachbarten Ecken mindestens eine die Bedingung r_3^* erfüllt. (Die Abschwächung der Geräumigkeitsbedingung stammt von K. Böröczky.) Es bedeutet soviel, daß der Radius von mindestens einer der benachbarten Stützkugeln des Mittelpunktsystems mindestens r_3^* ist. Dabei nennt man zwei Stützkugeln benachbart, wenn sie mindestens drei Punkte des Mittelpunktsystems gemeinsam haben. Offensichtlich sind die Mittelpunkte der benachbarten Stützkugeln benachbarte Ecken einer DV-Zelle.

Untersuchen wir jetzt eine Kugelpackung im dreidimensionalen euklidischen Raum mit Quasi-Geräumigkeitsbedingung r_3^*. Es bedeutet keine Beschränkung, wenn man ein Einheitskugelsystem betrachtet.

Satz. *Im dreidimensionalen euklidischen Raum sei ein Einheitskugelsystem* $\{K_i\}$, *das der Quasi-Geräumigkeitsbedingung* $r_3^* = \sqrt{2}$ *genügt, gegeben. Dann ist die Packungsdichte des Kugelsystems* $\{K_i\}$ *höchstens* $\frac{\pi}{\sqrt{18}}$. *Diese Abschätzung ist genau z.B. im Falle der dichtesten gitterförmigen Kugelpackung.*

Somit ist unter gewissen — nicht notwending gitterförmigen — Kugelpackungen die dichteste gitterförmige Kugelpackung eine der dichtesten. (Durch obigen Satz ist eine von K. Böröczky aufgestellte Vermutung bewiesen.)

Beim Beweis dieses Satzes spielt der von K. Böröczky eingeführte Begriff der *Grenzdichte* eine grundlegende Rolle.

Es seien im dreidimensionalen euklidischen Raum $Q_0Q_1Q_2Q_3$ ein Tetraeder und K eine Kugel mit dem Mittelpunkt Q_0. Bezeichne $\delta(Q_0, Q_1Q_2Q_3)$ die Dichte von K bezüglich des Tetraeders $Q_0Q_1Q_2Q_3$, die wie üblich definiert wird. Wir betrachten den Grenzwert $\delta(Q_0, Q_1, Z)$ von $\delta(Q_0, Q_1Q_2Q_3)$ für den Fall, daß Q_2 und Q_3 gegen einen gemeinsamen Punkt Z streben, wobei Z nicht auf der Geraden Q_0Q_1 liegt. Dabei gilt

$$\delta(Q_0, Q_1, Z) = \lim_{Q_2, Q_3 \to Z} \delta(Q_0, Q_1Q_2Q_3).$$

Bezeichne $R_a(X)$ einen Rotationskörper, der aus dem Gebiet X durch eine Drehung um die Achse a entsteht. Dann ist offenbar

$$\delta(Q_0, Q_1, Z) = \delta[R_{Q_0Q_1}(Q_0Q_1Z)].$$

Im Beweis kommt oft der Begriff des zwei- bzw. dreidimensionalen *Orthoschems* vor. Das zweidimensionale Orthoschem ist ein rechtwinkliges Dreieck $Q_0Q_1Q_Z$, wobei Q_0Q_1 senkrecht zu Q_1Q_Z ist. Das Tetraeder $Q_0Q_1Q_2Q_3$ ist ein Orthoschem, wenn die Kante Q_0Q_1 zu der Fläche $Q_1Q_2Q_3$ normal ist, wo $Q_1Q_2 \perp Q_2Q_3$ ist.

Für das Weitere benötigen wir die folgende Hilfssätze:

HILFSSATZ 1. *Es sei K eine Einheitskugel mit dem Mittelpunkt O. Ferner seien OO_1E_1 und OO_2E_2 je ein zweidimensionales Orthoschem, wo $OO_1 = OO_2 \geqq 1$ und $O_1E_1 < O_2E_2$ sind. Behauptung: $\delta(O, O_1, E_1) > \delta(O, O_2, E_2)$.*

HILFSSATZ 2. *Es seien $Q_0Q_1Q_2Q_3$ ein Tetraeder mit Volumen V und K eine Kugel mit Mittelpunkt Q_0, die die Ebene $Q_1Q_2Q_3$ nicht trifft. Ferner sei Z ein (innerer) Punkt der Strecke Q_2Q_3 und sei $\delta(Q_0, Q_1, Z)$ eine monotone Funktion von Q_2Z. Behauptung: Die Dichte $\delta(Q_0, Q_1Q_2Q_3)$ liegt zwischen den Grenzdichten $\delta(Q_0, Q_1, Q_2)$ und $\delta(Q_0, Q_1, Q_3)$.*

HILFSSATZ 3. *Es sei K eine Einheitskugel mit dem Mittelpunkt O. Ferner seien OO_1TA und $O\bar{O}_1\bar{T}\bar{A}$ zwei dreidimensionale Orthoscheme, wo $OO_1 \geqq O\bar{O}_1 \geqq 1, OT \geqq O\bar{T}$ und $OA \geqq O\bar{A}$ sind. Dann $\delta(O, O_1TA) \leqq \leqq \delta(O, \bar{O}_1\bar{T}\bar{A})$, und Gleichheit tritt dann und nur dann auf, wenn $OO_1 = O\bar{O}_1$, $OT = O\bar{T}$ und $OA = O\bar{A}$ gilt.*

(Hilfssätze 1-3 sind bei K. Böröczky [1], [2] zu finden.)

HILFSSATZ 4. *Es sei K eine Einheitskugel mit dem Mittelpunkt O. Ferner seien OO_1B und $O\bar{O}_1B$ je ein zweidimensionales Orthoschem, wo $OO_1 > O\bar{O}_1 \geqq 1$ ist (Abb. 1). Es folgt $\delta(O, O_1, B) \leqq \delta(O, \bar{O}_1, B)$.*

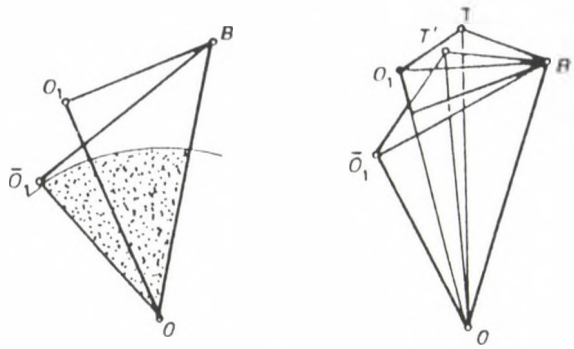


Abb. 1

BEWEIS. Es sei $\delta(O,O_1,B) = \lim\limits_{T\to B} \delta(O,O_1TB)$ und $\delta(O,\bar{O}_1,B) = \lim\limits_{T'\to B} \delta(O,\bar{O}_1T'B)$, wo OO_1TB und $O\bar{O}_1T'B$ je ein Orthoschem ist, sowie $OT = OT'$. Laut Hilfssatz 3 ist $\delta(O,O_1TB) < \delta(O,\bar{O}_1T'B)$. So gilt nach dem Grenzübergang

$$\delta(O,O_1,B) \leqq \delta(O,\bar{O}_1,B).$$

HILFSSATZ 5. *Es sei K eine Einheitskugel mit dem Mittelpunkt O. Ferner sei $O\bar{O}_1B$ ein zweidimensionales Orthoschem, wo $O\bar{O}_1 \geqq 1$ und $OB = \sqrt{\frac{3}{2}}$ ist (Abb. 2). Dann wird $\delta(O,\bar{O}_1,B) < d = \frac{\pi}{\sqrt{18}}$.*

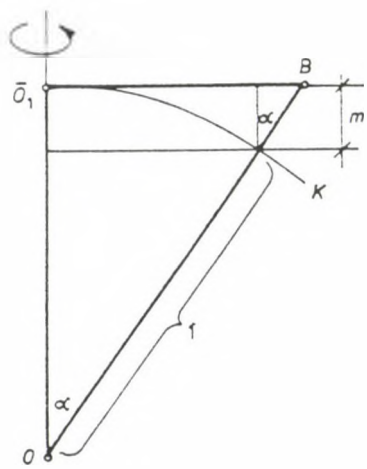


Abb. 2

BEWEIS. Drehen wir das rechtwinklige Dreieck $O\bar{O}_1B$ um die Achse $O\bar{O}_1$. Nach üblichen Definition von Grenzdichte und Dichte gilt

$$\delta(O,\bar{O}_1,B) = \delta[R_{O\bar{O}_1}(OO_1B)] = \frac{V_k(\alpha)}{V[R_{O\bar{O}_1}(O\bar{O}_1B)]},$$

wobei $V_k(\alpha)$ das Volumen eines Kugelsektors mit Winkel α ist ($\alpha = \sphericalangle(O\bar{O}_1, OB)$); und $V[R_{O\bar{O}_1}(O\bar{O}_1B)]$ das Volumen des Kegels $R_{O\bar{O}_1}(O\bar{O}_1B)$ ist. Dabei ergibt sich

$$V_k(\alpha) = \frac{2}{3}\pi r^2 m = \frac{2}{3}\pi\left(1 - \sqrt{\frac{2}{3}}\right),$$

wo $r = O\bar{O}_1 = 1$ und m ist die Höhe des Kugelsegments mit Winkel α, d.h.

$$m = \cos\alpha\left(\sqrt{\frac{3}{2}} - 1\right) = \left(1 - \sqrt{\frac{2}{3}}\right).$$

Weiters gilt

$$V[R_{OO_1}(O\bar{O}_1B)] = \frac{\pi}{3}r_k^2 m_k = \frac{\pi}{6},$$

wobei $r_k = \bar{O}B = \frac{\sqrt{2}}{2}$, $m_k = O\bar{O}_1 = 1$ sind. Deshalb ist $\delta(O, \bar{O}_1, B) = 4\left(1 - \sqrt{\frac{2}{3}}\right) = 0,734014\ldots < \frac{\pi}{\sqrt{18}}$. Also gilt $\delta(O, \bar{O}_1, B) < d$.

Wenden wir uns jetzt dem Beweis des Satzes zu.

Ohne Beschränkung der Allgemeinheit können wir uns auf eine gesättigte Kugelpackung $\{K_i\}$ beschränken, so daß im Raum leere Kugeln mit einem größeren Radius als $2\sqrt{2}$ nicht vorkommen können.

Wir betrachten jetzt die DV-Zellenzerlegung, die zum Kugelmittelpunktssystem gehört.Es sei K eine Kugel des $\{K_i\}$, deren Mittelpunkt O ist. Die DV-Zelle der Kugel K ist ein konvexes Polyeder P, denn es handelt sich um eine gesättigte Kugelpackung. Nach den trivialen Geräumigkeitsbedingungen betragen die Abstände der Flächenebenen, Kantengeraden bzw. Ecken des P vom Mittelpunkt O mindestens

$$r_1 = 1, \quad r_2 = \frac{2}{\sqrt{3}} \quad \text{bzw. } r_3 = \sqrt{\frac{3}{2}}.$$

Nach den Quasi-Geräumigkeitsbedingungen des Satzes ist der Abstand wenigstens einer der benachbarten Ecken (an der Kante) vom Mittelpunkt O das Minimum $r_3^* = \sqrt{2}$.

Wir betrachten die mit K konzentrischen Kugeln S_3^*, S_3 und S_2 vom Radius r_3^*, r_3 bzw. r_2 (Abb. 3).

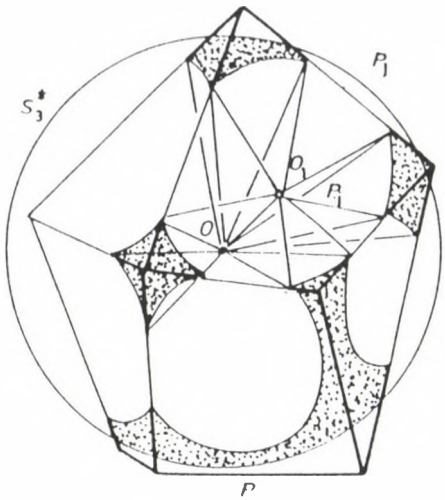


Abb. 3

Schälen wir mit der Kugel S_3^* das Polyeder P ab! Die Dichte der Kugel K in P ist höchstens so groß wie im enstandenen Körper PS_3^*, da der geschälte Teil leer ist ($P \cap S_3^* = PS_3^*$). PS_3^* ist im allgemeinen kein Polyeder mehr. Die Oberfläche des Körpers PS_3^* stammt einerseits von der Oberfläche der Kugel S_3^*, anderseits von der Oberfläche des Polyeders P.

Bezeichnen wir den Teil der Fläche PS_3^* auf der Fläche der Kugel S_3^* mit S und den Teil der Fläche PS_3^* auf einer Flächenebene des Polyeders P mit p_j (wo $j = 1, 2, \ldots, \ell$, falls ℓ Flächen des Polyeders P zur Entstehung von PS_3^* beitragen).

Zerlegen wir den Körper PS_3^* in ein kegelförmiges Gebilde mit Spitze O und dem Kugelflächen-Teil S als Grundfläche bzw. in die Kegel (Pyramiden) P_j mit Spitze O, die zu den Flächen p_j gehören.

In dem kegelförmigen Gebilde mit der Grundfläche S und der Spitze O beträgt die Dichte $\delta(S)$ der Kugel K:

$$\delta(S) = \left(\frac{r_1}{r_3^*}\right)^3 = \left(\frac{1}{\sqrt{2}}\right)^3 = 0,35355\ldots < d = \frac{\pi}{\sqrt{18}}.$$

Betrachten wir die Kegel P_j mit den Grundflächen p_j und der Spitze O, und untersuchen wir diesen die Dichte $S(P_j)$ der Kugel K. Der Fußpunkt der von O auf die Flächenebene von p_j gefällten Lote sei O_j. (O_j ist hier der Mittelpunkt des Kreises k_3^*, wo $k_3^* = S_3^* \cap p_j$ und $OO_j \geqq r_1 = 1$.) Wir unterscheiden zwei Fälle je nach dem, ob die Fläche p_j den Punkt O_j enthält oder nicht.

Wenn die Fläche p_j den Punkt O_j nicht enthält (O_j liegt außerhalb p_j), dann gibt es eine Seite (z.B. A_1A_k) von p_j, die p_j von O_j trennt (Abb. 4). (Sei $F \in p_j$ der O_i nächstliegende Punkt.)

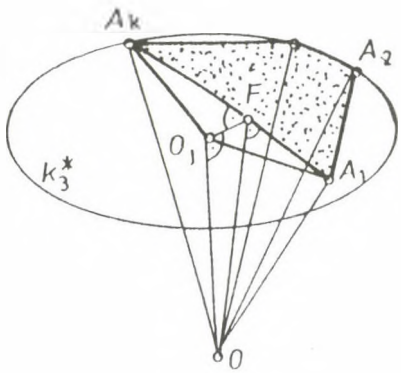


Abb. 4

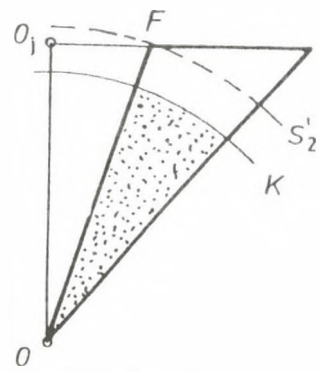


Abb. 5

Der Punkt F ist ein Punkt der Kante des Polyeders P, so gilt gemäß der Kantengeräumigkeits-Bedingung:

$$OF \geqq r_2 = \frac{2}{\sqrt{3}}.$$

Betrachten wir die mit der Kugel K konzentrische Kugel S_2' vom Radius OF und schneiden wir damit den Kegel P_j ab (Abb. 5). Im so abgestumpftem Kegel P_j' ist die Dichte der Kugel K größer als in P_j', da wir mit S_2 einen leeren Teil abgeschnitten haben. So es gilt:

$$\delta(P_j) < \delta(P_j') \leqq \frac{r_1^3}{r_2^3} = \frac{1}{(\frac{2}{\sqrt{3}})^3} = \frac{3\sqrt{3}}{8} = 0,64951 \ldots < d = \frac{\pi}{\sqrt{18}}.$$

Wir untersuchen im weiteren den Fall, daß die Fläche p_j den Punkt O_j enthält. (O_j liegt innerhalb p_j.) Wir unterscheiden weitere zwei Fälle je nachdem, $OO_j \geqq r_2$ bzw. $OO_j < r_2$.

(A) Wenn $OO_j \geqq r_2$; d.h., wenn die Grundebene p_j des Kegels P_j die Kugel S_2 vom Radius r_2 nicht durchschneidet, gilt die Dichtenabschätzung wie vorher

$$\delta(P_j) < \delta(\bar{P}_j) = \frac{r_1^3}{r_2^3} = 0,64951 \ldots < d = \frac{\pi}{\sqrt{18}}.$$

(Hier ist $\bar{P}_j$ ein durch eine Kugel S_2 abgestumpfter Kegel P_j.)

(B) Im Fall $OO_j < r_2$, also wenn die Grundebene p_j des Kegels P_j die Kugeln S_3^*, S_3 und S_2 vom Radius r_3^*, r_3 bzw. r_2 durchschneidet, dann seien $S_3^* \cap p_j = k_3^*$, $S_3 \cap p_j = k_3$ bzw. $S_2 \cap p_j = k_2$.

Es sei P_1 einer der Kegel P_j und sei O_1 die orthogonale Projektion des Kegelmittelpunktes O auf der Grundfläche p_j, wo $OO_1 \geqq 1$.

Bezeichnen wir die Ecken des Grundpolygons des Kegels P_1 mit A_i, die Fußpunkte der von O_1 auf der Seiten A_iA_{i+1} gefällten Lote mit T_i (Abb. 6).

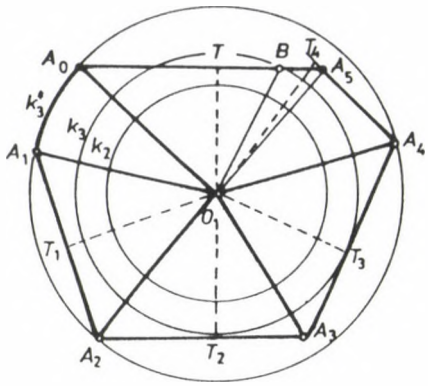


Abb. 6

Verbinden wir den Punkt O_1 mit den Ecken der Fläche p_1, und zerlegen wir zugleich den Kegel P_1 mittels Ebenen, die mit der Geraden OO_1 inzidieren, in Teilkegel. In diesen Teilkegeln werden wir die Dichte der Kugel K untersuchen.

Wenn man die Typen der Begrenzungselemente des Polygons P_1 in Betracht zieht, kann man die Teilkegel von P_1 folgendermaßen sortieren.

1. Das Begrenzungselement ist ein Kreisbogen (z.B. A_0A_1); der dazu gehörige Teilkegel ist *ein Kegel mit einer Kreissektor-Grundfläche* (z.B. $OO_1A_0A_1$).

2. Das Begrenzungselement ist eine Seitenstrecke die den Kreis k_3 nicht durchschneidet, höchstens berührt. Die dazu gehörenden Tetraeder $(OO_1A_iA_{i+1})$ kann man

   (a) im Fall $T \in A_iA_{i+1}$ zerlegen und zwar mittels der Ebenen OO_1T_i in *dreidimensionale Orthoscheme* (z.B. $OO_1T_1A_1$).

   (b) Im Fall $T_i \notin A_iA_{i+1}$ zerlegen wir die Tetraeder nicht weiter; (z.B. $OO_1A_4A_5$ ist ein sogenanntes *stumpfwinkligen Tetraeder* oder O-Tetraeder).

3. Das Begrenzungselement ist eine Seitenstrecke, die den Kreis k_3 durchschneidet. Die eine Ecke liegt auf dem Kreisbogen k_3 und die andere liegt im Bereich des Kreisringes $k_3^*k_3$ (z.B. A_0A_5). Wenn die andere Ecke nicht auf den Kreisbogen k_3 fällt, dann bezeichne B jenen Schnittpunkt der Strecke A_0A_5 mit dem Kreisbogen k_3, der von A_0 entfernter liegt, als A_5. Zerlegen wir das dazu gehörende Tetraeder (z.B. $OO_1A_0A_5$) mit einer eventuellen Ebene OO_1B in zwei weitere Tetraeder. Dabei ist

   a) eines ein Tetraeder von Typ OO_1BA_5 und

   b) das andere ein Tetraeder von Typ OO_1A_0B.

Wir untersuchen, wie sich die Kugeldichte bei allen möglichen Typen der im Laufe der Zerlegung entstandenen Kegel, Tetraeder und dreidimensionalen Orthoschemen gestaltet. Dabei werden die vorgeschriebenen Geräu-

migkeitsbedingungen in Betracht gezogen.

(1) Im Kegel mit Kreissektor-Grundfläche (z.B. $OO_1A_0A_1$) kann die Dichte der Kugel K auf Grund der Definition der Grenzdichte und des Hilfssatzes 1 folgendermaßen geschrieben werden:

$$\delta(O, O_1, A_0A_1) = \delta[R_{OO_1}(OO_1A_1)] = \delta(O, O_1, A_1) < \delta(O, O_1, B).$$

(2a) Betrachten wir in den Orthoschemen $OO_1T_iA_j$ ($j = i$ bzw. $j = i + 1$) — z.B. im $OO_1T_1A_1$ — die Dichte der Kugel K. Es sei E ein Punkt der Strecke T_iA_j. Laut des Hilfssatzes 1 ist $\delta(O, O_1, E)$ eine monotone abnehmende Funktion von T_iE. So gilt nach dem Hilfssatz 2:

$$\delta(O, O_1T_iA_j) < \delta(O, O_1, T_i).$$

Wieder Hilfssatz 1 angewandt gilt

$$\delta(O, O_1T_iA_j) < \delta(O, O_1, B).$$

(2b) Untersuchen wir die Kugeldichte im sogenannten stumpfwinkligen (oder O-) Tetraeder von Typ $OO_1A_4A_5$.

Es sei E ein Punkt der Strecke T_4A_4. Auf Grund des Hilfssatzes 1 ist die Grenzdichte eine monotone abnehmende Funktion von T_1E, so gilt mit Hilfssatz 2, und anschließend Hilfssatz 1

$$\delta(O, O_1A_4A_5) < \delta(O, O_1, A_5) < \delta(O, O_1B).$$

Betrachten wir nur die zweidimensionalen Orthoschemen OO_1B und $O\bar{O}_1B$, wo $OO_1 \geqq O\bar{O}_1 = 1$ ist. Laut Hilfssatz 4

$$\delta(O, O_1, B) \leqq \delta(O, \bar{O}_1, B),$$

da B mit k_3 inzident ist ($B \multimap k_3$), gilt $OB = \sqrt{\frac{3}{2}}$. Nach Berechnungen des Hilfssatz 5, ist

$$\delta(O, \bar{O}_1, B) < d = \frac{\pi}{\sqrt{18}}.$$

Dies mit dem oben unter (1), (2a) und (2b) gesagten, gilt

$$\delta(O, O_1A_0A_1) < d \text{ und } \delta(O, O_1A_iA_{i+1}) < d.$$

So ist bewiesen, daß in einem jeden Teilkegel der Gruppe (1) und (2) des Kegels P_1 die Dichte der Kugel K kleiner ist als d.

(3) Betrachten wir die Kugeldichte im Tetraeder des Types $OO_1A_0A_5$. Zerlegen wir es durch die Ebene OO_1B in zwei Tetraeder.

(a) Das Tetraeder OO_1BA_5 ist im wesentlichen ein Tetraeder von Typ (2b), weil $T \notin BA_5$ ist (siehe „O"-Tetraeder $OO_1A_4A_5$). Also gilt, wie oben bewiesen,

$$\delta(O, O_1BA_5) < \delta(O, O_1, B) < \delta(O, \bar{O}_1, B) < d.$$

(b) Schätzen wir die Kugeldichte K beim Tetraeder des Types OO_1AB, wo $A_0 = A \text{ —•— } k_3^*$, $B \text{ —o— } k_3$ ist. Drehen wir die Grundebene O_1AB des Tetraeders OO_1AB um die Achse AB in eine Tangentialebene der Kugel $K(AB \cap K = \emptyset)$. Bezeichnen wir den Berührungspunkt mit $\bar{O}_1$ (Abb. 7).

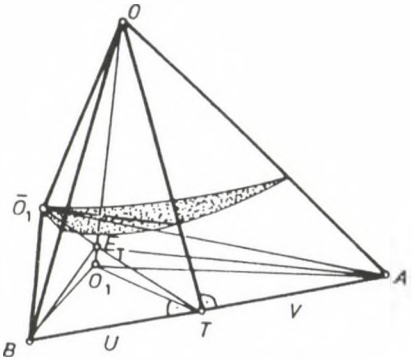


Abb. 7

(b1) *Behauptung*: $\delta(O, O_1AB) < \delta(O, \bar{O}_1AB)$.

Wir zerschneiden die Tetraeder OO_1AB und $O\bar{O}_1AB$ mit der Ebene $[OO_1T] = [O\bar{O}_1T]$ in je zwei Orthoscheme. Untersuchen wir in den so entstandenen Orthoschemen die Änderung der Kugeldichte.

Einerseits gilt für die Orthoscheme OO_1TB und OO_1TA wegen der Bedingung $OB < OA$ auf Grund des Hilfssatz 3:

$$\delta_B := \delta(O, O_1TB) > \delta(O, O_1TA) =: \delta_A.$$

Ebenso gilt nach der obigen Drehung

$$\bar{\delta}_B := \delta(O, \bar{O}_1TB) > \delta(O, \bar{O}_1TA) = \bar{\delta}_A.$$

Anderseits gilt für die Orthoscheme OO_1TB und $O\bar{O}_1TB$ sowie OO_1TA und $O\bar{O}_1TA$ ebenso wegen der Bedingung $OO_1 > O\bar{O}_1$ unter Anwendung des Hilfssatz 3:

$$\delta_B = \delta(O, O_1TB) < \delta(O, \bar{O}_1TB) = \bar{\delta}_B$$

$$\delta_A = \delta(O, O_1TA) < \delta(O, \bar{O}_1TA) = \bar{\delta}_A.$$

(Also wurde die Dichte von K nach der Drehung der Grundebene in den einzelnen Orthoschemen größer.)

Die Kugeldichte im ganzen Tetraeder ist das gewichtete Mittel der Dichten in den einzelnen Orthoschemen. Es bezeichne u und v die Länge der Strecken TB und TA, dann gilt

$$\frac{V_{OO_1TB}}{V_{OO_1TA}} = \frac{u}{v} = \frac{V_{O\bar{O}_1TB}}{V_{O\bar{O}_1TA}},$$

wo z.B. V_{OO_1TB} das Volumen des Tetraeders OO_1TB bezeichnet. So ist die Kugeldichte in den Tetraedern OO_1AB und $O\bar{O}_1AB$

$$\delta(O,O_1AB) := \delta = \frac{u\delta_B + v\delta_A}{u+v},$$

bzw.

$$\delta(O,\bar{O}_1AB) := \bar{\delta} = \frac{u\bar{\delta}_B + v\bar{\delta}_A}{u+v}.$$

Davon kommt offensichtlich

$$\delta = \delta(O,O_1AB) < \delta(O,\bar{O}_1AB) = \bar{\delta}.$$

(Also ist das Gewichtverhältnis der einzelnen Dichten nach dem Eindrehen unverändert geblieben, und die Dichte hat in Gesamtheit zugenommen.)

Danach drehen wir die Seitengerade AB des Grunddreiecks $\bar{O}_1AB$ um den Punkt A in die Lage, die den Kreis k_2 der Kugel S_2 berührt. Der Berührungspunkt sei $\bar{T}$, die Ecke B kommt sich am Kreis k_3 von A entfernend zum Punkt $\bar{B}$, und so geht das Dreieck $\bar{O}_1AB$ ins Dreieck $\bar{O}_1\bar{A}\bar{B}$ $(A=\bar{A})$ über (Abb. 8).

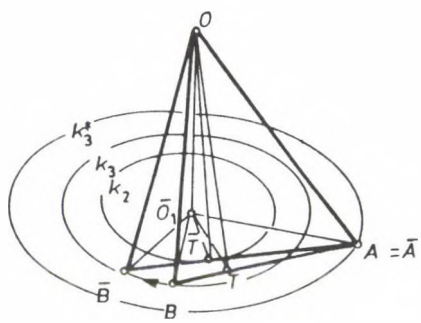


Abb. 8

(b2) *Behauptung*: $\delta(O,\bar{O}_1AB) < \delta(O,\bar{O}_1\bar{A}\bar{B})$.

Zerlegen wir die Tetraeder $O\bar{O}_1AB$ und $O\bar{O}_1\bar{A}\bar{B}$ mit der Ebene $[O\bar{O}_1T]$ bzw. $[O\bar{O}_1\bar{T}]$ in je zwei Orthoschemen. Wir untersuchen in den einzelnen Orthoschemen die Änderung der Dichte.

In den Orthoschemen $O\bar{O}_1T\bar{A}$ und $O\bar{O}_1\bar{T}A$, sowie $O\bar{O}_1TB$ und $O\bar{O}_1\bar{T}\bar{B}$ gilt wegen der Ungleichung $OT > O\bar{T}$ im Sinne des Hilfssatzes 3:

$$\bar{\delta}_A = (O,\bar{O}_1TA) < \delta(O,\bar{O}_1\bar{T}\bar{A}) = \bar{\delta}_{\bar{A}},$$

sowie

$$\bar{\delta} = (O,\bar{O}_1TB) < \delta(O,\bar{O}_1\bar{T}\bar{B}) = \bar{\delta}_{\bar{B}}.$$

Das Verhältnis des Volumens bei den einzelnen Orthoschemen beträgt vor und nach der Drehung:

$$\frac{V_{O\bar{O}_1TB}}{V_{O\bar{O}_1TA}} = \frac{x}{x+y},$$

wo $x = TB = TB'$, $y = AB'$ sind, bzw.

$$\frac{V_{O\bar{O}_1\bar{T}\bar{B}}}{V_{O\bar{O}_1\bar{T}\bar{A}}} = \frac{x'}{x'+y'},$$

wo $x' = \bar{T}\bar{B} = \bar{T}\bar{B}'$, $y = \bar{A}\bar{B}'$ sind (B' bzw. $\bar{B}'$ bezeichnet die zweiten Schnittpunkte der Strecken AB bzw. $\bar{A}\bar{B}$ mit Kreis k_3). So beträgt die Kugeldichte in den Tetraedern $O\bar{O}_1AB$ und $O\bar{O}_1\bar{A}\bar{B}$:

$$\delta(O, \bar{O}_1AB) = \bar{\delta} = \frac{x\bar{\delta}_B + (x+y)\bar{\delta}_A}{2x+y}$$

$$\delta(O, \bar{O}_1\bar{A}\bar{B}) := \bar{\bar{\delta}} = \frac{x'\bar{\delta}_B + (x'+y')\bar{\delta}_{\bar{A}}}{2x'+y'}.$$

Untersuchen wir die Änderung des Gewichtverhältnisses! (Abb. 9.) Offensichtlich $2x' > 2x$ und $0 < y' < y$, denn in den Dreiecken $A\bar{O}_1B'$ und $\bar{A}\bar{O}_1\bar{B}'$ für die Winkel gilt:

$$A\bar{O}_1B'\sphericalangle > \bar{A}\bar{O}_1\bar{B}'\sphericalangle, \quad \text{so ist } AB' > \bar{A}\bar{B}'.$$

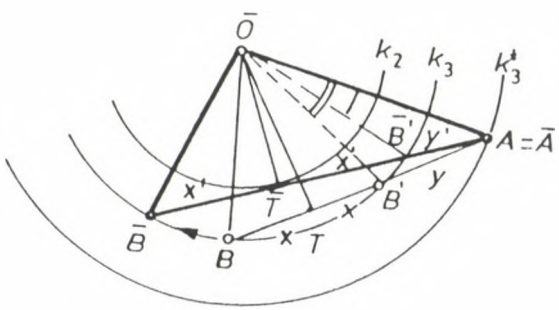


Abb. 9

Deshalb

$$\frac{x'}{y'} > \frac{x}{y} \quad \text{d.h.} \quad \frac{x'}{2x'+y'} > \frac{x}{2x+y}.$$

Das Gewichtverhältnis des Orthoschemes mit Ecke „B" hat bei der Drehung zugenommen.

Drücken wir diese Ungleichungen durch gewichtete Mittelwerte aus:

$$\bar{\delta} = \delta(O, \bar{O}_1 AB) = \frac{x\bar{\delta}_B + (x+y)\bar{\delta}_A}{2x+y} < \frac{x}{2x+y}\bar{\delta}_B + \frac{x+y}{2x+y}\bar{\delta}_{\bar{A}} =$$

$$= \frac{x'}{2x'+y'}\bar{\delta}_B + \left(\frac{x'}{2x'+y'} - \frac{x}{2x+y}\right)(\bar{\delta}_{\bar{A}} - \bar{\delta}_B) + \frac{x'+y'}{2x'+y'}\bar{\delta}_{\bar{A}} <$$

$$< \frac{x'}{2x'+y'}\bar{\delta}_B + \frac{x'+y'}{2x'+y'}\bar{\delta}_{\bar{A}} = \delta(O, \bar{O}_1 \bar{A}\bar{B}) = \bar{\bar{\delta}}$$

Im Tetraeder $O\bar{O}_1\bar{A}\bar{B}$ beträgt die Dichte

$$(O, \bar{O}_1\bar{A}\bar{B}) = d = \frac{\pi}{\sqrt{18}}.$$

Das um die Einheitskugel geschrieben Rhombendodekaeder kann in 48 Stücke zerschnitten werden, welche mit dem Tetraeder $O, \bar{O}_1\bar{A}\bar{B}$ kongruent sind. Wenn man die Flächen- und Winkelverhältnisse des Tetraeders $O\bar{O}_1\bar{A}\bar{B}$ in Betracht zieht, sieht man daß nur das Rhombendodekaeder in Teile zerlegt werden kann, die mit dem Tetraeder $O\bar{O}_1\bar{A}\bar{B}$ kongruent sind.

Die DV-Zellen der dichtesten gitterförmigen Kugelpackung sind aber Rhombendodekaeder.

Die Behauptungen (b1) und (b2) zusammenfassend gilt:

$$\delta(O, O_1AB) < \delta(O, \bar{O}_1AB) < \delta(O, \bar{O}_1\bar{A}\bar{B}) = d$$

d.h. es ist

$$\delta < \bar{\delta} < \bar{\bar{\delta}} = d.$$

Es ist also bewiesen, daß die Dichte der Kugel K in jedem Teiltetraeder der Gruppe (3) des Kegels P_1 höchstens $d = \frac{\pi}{\sqrt{18}}$ ist.

Wir haben bewiesen, daß die Dichte der Kugel K in einem beliebigen Kegel P_j, im Konvexkörper PS_3^* und so im Polyeder P d.h. in einer beliebigen DV-Zelle des Einheitskugelsystems $\{K_i\}$, das den Quasi-Geräumigkeitsbedingungen $r_3^* = \sqrt{2}$ genügt, höchstens d ist.

Diese Dichte kann aber dann vorkommen, wenn Polyeder P in mit dem Tetraeder $O\bar{O}_1\bar{A}\bar{B}$ kongruente Teile zerlegt werden kann, d.h. wenn Polyeder P ein Rhombendodekaeder ist.

Die dichteste gitterförmige Kugelpackung ist auch unter den Kugelpackungen mit Quasi-Geräumigkeitsbedingung $r_3^* = \sqrt{2}$; es ist die dichteste. D.h. unter den betrachten nicht unbedingt gitterförmigen Kugelpackungen ist es gelungen, die dichteste gitterförmige Kugelpackung als eine der dichtesten anzugeben.

LITERATURVERZEICHNIS

[1] BÖRÖCZKY, K., Gömbkitöltések állandó görbületű terekben [Sphere packing in spaces of constant curvature] II, *Mat. Lapok* **26** (1975), 67–90 (in Hungarian). *MR* **58** #24015

[2] BÖRÖCZKY, K. und FLORIAN, A., Über die dichteste Kugelpackung in hyperbolischen Raum, *Acta Math. Acad. Sci. Hungar.* **15** (1964), 237–245. *MR* **28** #3369

[3] FEJES TÓTH, L., *Lagerungen in der Ebene, auf der Kugel und im Raum,* Zweite verbesserte und erweiterte Auflage, Die Grundlehren des math. Wissenschaften, Band 65, Springer-Verlag, Berlin–New York, 1972. *MR* **50** # 5603

[4] FEJES TÓTH, L., *Reguläre Figuren*, Akadémiai Kiadó, Budapest, 1965. *MR* **30** # 3408

[5] HOLLAI, M., Das dichteste gitterförmige ρ-System der Kugeln, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **24** (1981), 157–180. *MR* **88b**: 52029

[6] MOLNÁR, J., Körelhelyezések állandó görbületű felületeken, *Magyar Tud. Akad. Mat. Fiz. Oszt. Közl.* **12** (1962), 223–263. *MR* **28** # 1535

[7] MOLNÁR, J., Kreislagerungen auf Flächen konstanter Krümmung, *Math. Ann.* **158** (1965), 365–376. *MR* **31** # 2663

[8] MOLNÁR, J., On the ρ-system of unit circles, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **20** (1977), 195–203. *MR* **58** # 12733

[9] MOLNÁR, J., Packing of congruent spheres in a strip, *Acta Math. Acad. Sci. Hungar.* **31** (1978), 173–183. *MR* **58** # 7406

(Eingegangen am 26. August 1985.)

YBL MIKLÓS FŐISKOLA
THÖKÖLY ÚT 74
H–1148 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 447–452

MATRIX EQUATION IN RADICALS

M. ASLAM and A.M. ZAIDI

Abstract

Let R_n be a complete ring of $n \times n$ matrices over a ring R. R. E. Propes [5] has given necessary conditions under which a certain radical class $\mathcal{P}$ satisfies the matrix equation $\mathcal{P}(R_n) = (\mathcal{P}(R))_n$.

In this paper, necessary conditions are found for a certain class of rings under which lower radical and upper radical class satisfy the matrix equation. Moreover, the idea of a matrix equation is extended to the sum of radical classes.

Introduction

Throughout this paper, we shall work within the class of associative rings. R. E. Propes [5] has discussed necessary conditions under which a certain radical class $\mathcal{P}$ satisfies the matrix equation $\mathcal{P}(R_n) = (\mathcal{P}(R_n))_n$, where R_n denotes the complete ring of $n \times n$ matrices over a ring R. In the present paper, we will investigate conditions under which matrix property of certain classes $\mathcal{M}$ and $\mathcal{M}^*$ of rings is carried over to lower radical class $L\mathcal{M}$ or the upper radical class $U\mathcal{M}^*$. Moreover we extend the idea of matrix equation of radical classes to the sum of radical classes (c.f. [4]). $\mathcal{W}$ will denote a universal class of all associative rings and $I \leqq R, J \lneqq R$ denote ideals of R, but $J \neq R$.

1. Matrix equation in lower radicals

Let $\mathcal{M}$ be a subclass of $\mathcal{W}$ and let $\mathcal{M}_o$ be the homomorphic closure of $\mathcal{M}$. We define the following classes from a given ring R:

$$H R = \text{ set of all homomorphic images of } R;$$

$$D_1(R) = \{I : I \leqq R\}.$$

1980 *Mathematics Subject Classification.* Primary 16A21.

Key words and phrases. Ring, radical classes, matrix equation, homomorphic closure, hereditary, homomorphically close, embed, upper radical, lower radical, regular class of rings, Baer lower radical, nilpotent ring, Brown–McCoy radical, sum of two radical classes.

Akadémiai Kiadó, Budapest

Inductively define

$$D_{n+1}(R) = \{I : I \text{ is an ideal of some ring in } D_n(R)\}$$
$$D(R) = \bigcup D_n(R),\ n = 1, 2, \ldots .$$

Then by [2], $L\mathcal{M} = \{R \in \mathcal{W} : D(R/I) \cap \mathcal{M}_o \neq 0,\ \forall\ I \lneqq R\}$ is the Lee construction of lower radical class determined by $\mathcal{M}$. Suliński et alia [7] gave the following characterization of the lower radical class determined by homomorphically closed class $\mathcal{M}$ of rings. Let $\mathcal{M} = \mathcal{M}_1$ and $\mathcal{M}_m$ has been defined for $m < n$. Define

$$\mathcal{M}_n = \{R \in \mathcal{W} : D_1(R/I) \cap \mathcal{M}_m \neq 0 \text{ for some } m < n,\ \forall\ I \lneqq R\}.$$

Then $L\mathcal{M} = \bigcup \mathcal{M}_n$, $n = 1, 2, 3 \ldots$ is the lower radical class determined by $\mathcal{M}$. Observe that $0 \in \mathcal{M}$, whenever $\mathcal{M}$ is homomorphically closed. Further, $\mathcal{SP}$ will denote the semisimple class of the radical class $\mathcal{P}$. For more details of radical theory we refer to [8] and [9].

LEMMA 1.1. *If $\mathcal{M}$ is a class of rings such that $R \in \mathcal{M}$ implies that $R_n \in \mathcal{M}$ then $\mathcal{M}_o$ has this property on rings, too.*

PROOF. Let $R \in \mathcal{M}_o$, then $R = S/I$ for some $S \in \mathcal{M}$ and $R_n = (S/I)_n \cong S_n/I_n$. Since $S \in \mathcal{M}$ implies that $S_n \in \mathcal{M}$, therefore $S_n/I_n \in \mathcal{M}_o$ and hence $R_n \in \mathcal{M}_o$. This completes the proof.

THEOREM 1.2. *If $\mathcal{M}$ is a class of rings such that $R \in \mathcal{M}$ implies that $R_n \in \mathcal{M}$, then $L\mathcal{M}$ has this property.*

PROOF. We will show that if $R_n \notin \mathcal{M}$ implies that $R \notin \mathcal{M}$, then $L\mathcal{M} = \mathcal{P}$ has this property. Let $R_n \notin L\mathcal{M}$ then by Snider [6, Lemma 7] $\mathcal{P}(R_n) = I_n$ for some $I \lneqq R$. We will show that $D(R/I) \bigcap \mathcal{M}_o = 0$. Let $J/I \in D(R/I) \bigcap \mathcal{M}_o$; assume that $0 \neq J/I$. Now we will have sequence $J_1/I, J_2/I, \ldots, J_m/I$ of subrings R/I such that $J/I \leqq J_1/I \ldots \leqq R/I$. This implies that $(J/I)_n \leqq \leqq (J_1/I)_n \ldots \leqq (J_m/I)_n \leqq (R/I)_n$, since $(R/I)_n \cong R_n/I_n$ and $R_n/I_n \in s\mathcal{P}$, and $s\mathcal{P}$ is hereditary, therefore $(J/I)_n \in s\mathcal{P}$. (Here $s\mathcal{P}$ denotes the semisimple class of $\mathcal{P}$.) By $\mathcal{P} \cap s\mathcal{P} = 0$ and $\mathcal{M}_o \subseteq \mathcal{P}$, it follows that $(J/I)_n \notin \mathcal{M}_o$. By Lemma 1.1, $J/I \notin \mathcal{M}_o$ which contradicts the fact that $J/I \in \mathcal{M}_o$ and hence $J/I = 0$. This shows that $D(R/I) \cap \mathcal{M}_o = 0$ for $I \lneqq R$ and consequently $R \notin L\mathcal{M}$.

LEMMA 1.3. *If $\mathcal{M}$ is a homomorphically closed class of rings (not necessarily all with unity) and $R_n \in \mathcal{M}$ implies that $R \in \mathcal{M}$, then $\mathcal{M}_2$ has this property on rings with unity.*

PROOF. Let R be a ring with unity, and let $R \notin \mathcal{M}_2$. Then there exists $0 \neq R/I$ such that $D_1(R/I) \cap \mathcal{M}_1 = 0$ and hence $R/I \notin \mathcal{M}_1 = \mathcal{M}$. This implies that $(R/I)_n \notin \mathcal{M}$. To show that $D_1((R/I)_n) \cap \mathcal{M}_1 = 0$ let $0 \neq L_n \in \in D_1((R/I)_n)$. Since R is a ring with unity, therefore by [2, p. 38], $L_n = = (K/I)_n$ for some $K/I \leqq R/I$ and hence $K/I \notin \mathcal{M}$ (by $D_1(R/I) \cap \mathcal{M}_1 = 0$).

This implies that $(K/I)_n \notin \mathcal{M}_1$ or $L_n \notin \mathcal{M}_1$. As L_n is an arbitrary non-zero ideal of $(R/I)_n$ such that $L_n \notin \mathcal{M}_1$, this shows that $D_1((R/I)_n) \cap \mathcal{M}_1 = 0$ and consequently $D_1(R_n/I_n) \cap \mathcal{M}_1 = 0$. This implies that $R_n \notin \mathcal{M}_2$. Now $R \notin \mathcal{M}_2$ implies that $R_n \notin \mathcal{M}_2$ or $R_n \in \mathcal{M}_2$ implies that $R \in \mathcal{M}_2$. This completes the proof.

LEMMA 1.4. *If $\mathcal{M}$ is a homomorphically closed class of rings and $R_n \in \in \mathcal{M}$ implies that $R \in \mathcal{M}$, then for each n, $\mathcal{M}_n$ has this property on rings with unity.*

PROOF. Straightforward by induction.

THEOREM 1.5. *If $\mathcal{M}$ is a homomorphically closed class of rings and $R_n \in \mathcal{M}$ implies that $R \in \mathcal{M}$, then $L\mathcal{M}$ has this property on rings with unity.*

PROOF. Let R be a ring with unity, and suppose that $R \notin L\mathcal{M} = \bigcup \mathcal{M}_n$. Then $R \notin \mathcal{M}_n$ for all n. By Lemma 1.4, $R_n \notin \mathcal{M}_n$ for all n. This implies that $R_n \notin \bigcup \mathcal{M}_n$ or $R_n \notin L\mathcal{M}$, which completes the proof.

LEMMA 1.6. *Let $\mathcal{P}$ be a hereditary radical class which satisfies the matrix equation on rings with unity. Then $\mathcal{P}$ also satisfies the matrix on the ring without unity.*

PROOF. Suppose $R \in \mathcal{W}$ has no unity element. Embed R into a ring S with unity. Since $\mathcal{P}$ is hereditary, therefore by [9, Theorem 13.1] $\mathcal{P}(R) = R \cap \cap \mathcal{P}(S)$, and hence

$$(1) \qquad (\mathcal{P}(R))_n = (R \cap \mathcal{P}(S))_n = R_n \cap (\mathcal{P}(S))_n = R_n \cap (\mathcal{P}(S_n)).$$

Since $\mathcal{P}$ is a hereditary and $R_n \leqq S_n$, therefore $\mathcal{P}(R_n) = R_n \cap (\mathcal{P}(S_n))$. This proves that $\mathcal{P}(R_n) = (\mathcal{P}(R))_n$ (by (1)).

By [5] and Theorems 1.5, 1.2 and Lemma 1.6, we obtain the following

COROLLARY 1.7. *If $\mathcal{M}$ is a class of rings which is homomorphically closed, hereditary and $R \in \mathcal{M}$ if and only if $R_n \in \mathcal{M}$. Then $\mathcal{LM}$ satisfies the matrix equation.*

Remark that Corollary 1.7 proves the well-known classical result that the Baer lower radical class of all nilpotent rings satisfies the matrix equation.

2. Matrix equation in upper radicals

A subclass $\mathcal{M}^*$ of $\mathcal{W}$ is said to be regular if $0 \in \mathcal{M}^*$ and $R \in \mathcal{M}^*$ implies that $HI \cap \mathcal{M}^* \neq 0$, $\forall\, 0 \neq I \in D_1(R)$. By [9, Theorem 7.2], $U\mathcal{M}^* = \{R \in \in \mathcal{W} : HR \cap \mathcal{M}^* = 0\}$ is the upper radical determined by $\mathcal{M}^*$.

THEOREM 2.1. *If $\mathcal{M}^*$ is a regular class of rings such that $R \in \mathcal{M}^*$ if and only if $R_n \in \mathcal{M}^*$. Then $U\mathcal{M}^* = \mathcal{P}$ satisfies the matrix equation on rings with unity.*

PROOF. Let R be a ring with unity and $R \in U\mathcal{M}^*$. Suppose $R_n \notin U\mathcal{M}^*$, then we have $0 \neq R_n/J \notin \mathcal{M}^*$. Since R is a ring with unity, therefore $J = I_n$ for some $I \leqq R$ and hence $R_n/J = R_n/I_n$. By $(R/I)_n \cong R_n/I_n$, it follows that we have $R/I \in \mathcal{M}^*$ and hence $R \notin U\mathcal{M}^*$ which is a contradiction. This proves that $R_n \in U\mathcal{M}^*$. For the converse, let $R_n \in U\mathcal{M}^*$. Suppose that $R \notin U\mathcal{M}^*$. Then we have $0 \neq R/I$ such that $R/I \in \mathcal{M}^*$. This implies that $(R/I)_n \in \mathcal{M}^*$ or $R_n/I_n \in \mathcal{M}^*$. It follows that $R_n \notin U\mathcal{M}^*$ which is a contradiction. By [5], $\mathcal{P}(R_n) = (\mathcal{P}(R))_n$.

Let $\mathcal{M}^*$ be the class of all simple rings with unity and $P = U\mathcal{M}^*$ be its upper radical class, known as Brown–McCoy radical class. It is easy to see that $R \in \mathcal{M}^* \iff R_n \in \mathcal{M}^*$ and hence by the above theorem $\mathcal{P}$ satisfies the matrix equation. This proves the following classical result.

COROLLARY 2.2 (see [1], [8, page 171 Theorem 38.7]). *The Brown–McCoy radical class satisfies the matrix equation.*

Lemma 1. 6 and the above theorem lead to the following

COROLLARY 2.3. *If $\mathcal{M}^*$ is a regular class of rings such that $R \in \mathcal{M}^*$ if and only if $R_n \in \mathcal{M}^*$; and $U\mathcal{M}^* = \mathcal{P}$ is hereditary, then $\mathcal{P}$ satisfies the matrix equation.*

A class $\mathcal{M}^*$ of rings is a special class of rings (see for instance [8, page 68]), in the sense of Andrunakievich if $\mathcal{M}^*$ is hereditary, consists of prime rings and is closed under essential extension, that is if I is an essential (or large) ideal of a ring R and $I \in \mathcal{M}^*$ then $R \in \mathcal{M}^*$. The upper radical $\mathcal{U}\mathcal{M}^*$ of the special class $\mathcal{M}^*$ is called special radical. It is well-known that special radicals are always hereditary (see for instance, [8]) and by Corollary 2.3, we have the following

COROLLARY 2.4. *If $\mathcal{M}^*$ is a special class of rings such that $R \in \mathcal{M}^*$ if and only if $R_n \in \mathcal{M}^*$, then the special radical $\mathcal{P} = \mathcal{U}\mathcal{M}^*$ satisfies the matrix equation.*

Remark that one can generalize the above corollary by taking $\mathcal{M}^*$ as weakly special class (for the definition see [8, page 66]) instead of special class. The upper radical $\mathcal{U}\mathcal{M}^* = \mathcal{P}$ generated by weakly special class is known as super nilpotent radical, which is always special (e.g. [8]). This leads to the following

COROLLARY 2.5. *If $\mathcal{M}^*$ is a weakly special class of rings such that $R \in \in \mathcal{M}^*$ if and only if $R_n \in \mathcal{M}^*$, then the super nilpotent radical $\mathcal{P} = \mathcal{U}\mathcal{M}^*$ satisfies matrix equation.*

3. Matrix equation in sum of two radical classes

If $\mathcal{P}_1$ and $\mathcal{P}_2$ are two radical classes, then the sum is defined as $\mathcal{P}_1 + \mathcal{P}_2 = \{R \in W : \mathcal{P}_1(R) + \mathcal{P}_2(R) = R\}$. In [4] it was shown that $\mathcal{P}_1(R) + \mathcal{P}_2(R)$ is the largest $(\mathcal{P}_1 + \mathcal{P}_2)$-ideal. Hence we can write $\mathcal{P}_1(R) + \mathcal{P}_2(R) = (\mathcal{P}_1 + \mathcal{P}_2)(R)$. We say that $\mathcal{P}_1 + \mathcal{P}_2$ satisfies the matrix equation if $(\mathcal{P}_1 + \mathcal{P}_2)(R_n) = ((\mathcal{P}_1 + \mathcal{P}_2)(R))_n$, $\forall\, R \in W$.

We shall frequently use the following

LEMMA 3.1 [2]. *Let R be a ring and I, J be ideals of R; then $(I + J)_n = I_n + J_n$.*

THEOREM 3.2. *If $\mathcal{P}_1$ and $\mathcal{P}_2$ are radical classes and R is a ring, then $(\mathcal{P}_1 + \mathcal{P}_2)(R_n) = I_n$ for some $I \trianglelefteq R$.*

PROOF. This follows from Lemma 3.1 and by [6, Lemma 7].

THEOREM 3.3. *If $\mathcal{P}_1$ and $\mathcal{P}_2$ are radical classes of rings satisfying the matrix equation, then $\mathcal{P}_1 + \mathcal{P}_2$ also satisfies the matrix equation.*

PROOF. This is obvious from Lemma 3.1.

THEOREM 3.4. *Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be radical classes and R be a ring. Then the following statements are equivalent:*

$$\text{(i)} \quad R \in (\mathcal{P}_1 + \mathcal{P}_2) \Rightarrow R_n \in (\mathcal{P}_1 + \mathcal{P}_2);$$

$$\text{(ii)} \quad ((\mathcal{P}_1 + \mathcal{P}_2)(R))_n \subseteq (\mathcal{P}_1 + \mathcal{P}_2)(R_n).$$

PROOF. This is similar to [5, Theorem 1]; use the fact that $(\mathcal{P}_1 + \mathcal{P}_2)(R_n)$ ist the largest $(\mathcal{P}_1 + \mathcal{P}_2)$-ideal of R_n (c.f. [4]).

THEOREM 3.5. *Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be radical classes and R be a ring. Then the following statements are equivalent*

$$\text{(i)} \quad R_n \in (\mathcal{P}_1 + \mathcal{P}_2) \Rightarrow R \in (\mathcal{P}_1 + \mathcal{P}_2);$$

$$\text{(ii)} \quad (\mathcal{P}_1 + \mathcal{P}_2)(R_n) \subseteq ((\mathcal{P}_1 + \mathcal{P}_2)(R))_n.$$

PROOF. This is obtained from [5, Theorem 2] by using Theorem 3.2 instead of Lemma 7 of [6].

COROLLARY 3.6. *If $\mathcal{P}_1$ and $\mathcal{P}_2$ are radical classes, then $\mathcal{P}_1 + \mathcal{P}_2$ satisfies the matrix equation on a ring R, if and only if $R \in \mathcal{P}_1 + \mathcal{P}_2 \iff R_n \in \mathcal{P}_1 + \mathcal{P}_2$.*

THEOREM 3.7. *If $\mathcal{P}_1 + \mathcal{P}_2$ is hereditary sum of two radical classes $\mathcal{P}_1$ and $\mathcal{P}_2$ such that $\mathcal{P}_1 + \mathcal{P}_2$ satisfies the matrix equation on rings with unity, then it also satisfies on rings without unity.*

PROOF. This is similar to that of Lemma 1.6. Use [4, Proposition 6] instead of [9, Theorem 13.1].

ACKNOWLEDGEMENTS. We acknowledge Dr. A. B. Thaheem for making many valuable suggestions. The authors also thank the referee for his useful comments to improve the paper.

REFERENCES

[1] BROWN, B. and MCCOY, N. H., The radical of a ring, *Duke Math. J.* **15** (1948), 495–499. *MR* **10**-6

[2] BURTON, D. M., *A first course in rings and ideals*, Addison–Wesley Publ. Co., Reading, Mass.–London–Don Mills, Ont., 1970. *MR* **41** #3509

[3] LEE, Y., On the construction of lower radical properties, *Pacific J. Math.* **28** (1969), 393–395. *MR* **39** #1492

[4] LI, Y. L. and PROPES, R. E., The sum of two radical classes, *Kyungpook Math. J.* **13** (1973), 81–86. *MR* **48** #325

[5] PROPES, R. E., The radical equation $P(A_n)=(P(A))_n$, *Proc. Edinburgh Math. Soc.* (2) **19** (1974/75), 257–259. *MR* **52** #468

[6] SNIDER, R., Lattices of radicals, *Pacific J. Math.* **40** (1972), 207–220. *MR* **46** #7290

[7] SULIŃSKI, A., ANDERSON, T. and DIVINSKY, N., Lower radical properties for associative rings and alternative rings, *J. London Math. Soc.* **41** (1966), 417–424. *MR* **33** #4095

[8] SZÁSZ, F. A., *Radicals of rings*, Wiley, Chichester, 1981. *MR* **84a**:16012

[9] WIEGANDT, R., *Radical and semisimple classes of rings*, Queen's Papers in Pure and Applied Mathematics, No. 37, Queen's University, Kingston, Ont., 1974. *MR* **50** #2227

(Received September 20, 1987)

PRINCIPAL GORDON COLLEGE
RAWALPINDI
PAKISTAN

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 453–455

ON THE PRODUCT OF k- AND l-SPACES

H. RENDER

0. Introduction

Lambrinos raised the question whether the product of a locally bounded space X and an arbitrary l-space Y is an l-space. We show that this is not true even if X is a compact space. Thus the category of all l-spaces is not convenient in the sense of [9]. On the other side we show that the product $X \times Y$ is an l-space if X is a basic locally compact space extending a result in [3, Proposition 1.7]. Another byproduct of our investigations is the following result: the product $T \times Y$ of two k-spaces T, Y is a k-space iff the exponential law is valid for the triple (T, Y, Z) with Z arbitrary. A similar result is valid for k_3-spaces where Z is an arbitrary regular space.

1. The results

Let Y, Z be topological spaces and α be an arbitrary system of subsets of Y. Then $C_\alpha(Y, Z)$ denotes the set of all functions whose restrictions on each set $A \in \alpha$ are continuous and $C(Y, Z)$ the set of all continuous functions. We call Y an α-space (resp. α_3-space) if $C_\alpha(Y, Z) = C(Y, Z)$ holds for all topological (resp. regular) spaces Z. As in [4] l denotes the system of all closures of bounded subsets. For unexplained terminology we refer to [4].

The set-open topology τ_α on $C_\alpha(Y, Z)$ is generated by the sets $[A, V] := \{f \in C_\alpha(Y, Z) : f(A) \subset V\}$ with $A \in \alpha$ and V open. Let $(f_i)_{i \in I}$ be a net in $C_\alpha(Y, Z)$. We say that $(f_i)_i$ *converges α-continuously* to $f \in C_\alpha(Y, Z)$ provided that for every $A \in \alpha$, for every $y \in Y$ and for every net $(y_j)_{j \in J}$ in A converging to $y \in A$ the net $(f_i(y_j))$ converges to $f(y)$. The following proposition can be seen as an improvement of Corollary 3.1.4 (a) in [6] and it shows that τ_α is a splitting topology, cf. Theorem 3.1.2 and Theorem 2.5.2 in [6].

1980 *Mathematics Subject Classification.* Primary 54D50; Secondary 54D99.
Key words and phrases. l-spaces, k-spaces, locally bounded spaces.

Akadémiai Kiadó, Budapest

PROPOSITION 1. *Let Y, Z be topological spaces and α be a system of compact subsets of Y. If a net in $C_\alpha(Y,Z)$ converges α-continuously to $f \in C(Y,Z)$ then it converges to f with respect to τ_α.*

PROOF. We show that f_i is in $[A,V]$ for almost all $i \in I$. Assume the contrary. Then there exists infinitely many $i \in I$ and $y_i \in A$ such that $f_i(y_i) \notin V$. Furthermore there exists a subnet $(y_j)_{j\in J}$ converging to some y in the compact space A. Clearly $(f_j)_{j\in J}$ converges α-continuously to f; therefore we obtain $f_j(y_j) \to f(y) \in V$, a contradiction.

Suppose now that the evaluation $e: C_\alpha(Y,Z)\times A \to Z$ defined by $e(f,x) := f(x)$ is continuous for τ_α and for every $A \in \alpha$. Then the following converse of Proposition 1 is obvious:

(1) If $(f_i)_i$ converges to $f \in C_\alpha(Y,Z)$ then $(f_i)_i$ converges α-continuously.

(1) is satisfied for the system k of all compact sets if each $K \in k$ is contained in a basic locally compact space *or* Z is regular or Hausdorff, see [5]. More generally (1) is satisfied for a so-called hereditarily closed network α (for definition see [6, p. 5]) if Y *or* Z is a regular space. For example, if cs is the system of all convergent sequences in a metric space X then Theorem 2 shows that the exponential law is valid for all metric spaces Y, Z with respect to the topology τ_{cs} which is in general strictly coarser than τ_k, cf. [2]. If β, α are families of subsets then $\beta \otimes \alpha$ denotes the system of all sets $B \times A$ with $B \in \beta$, $A \in \alpha$.

THEOREM 2. *Let T, Y, Z be topological spaces, α be a family of compact subsets of Y and β be a family with $\cup_{B\in\beta} B = T$. Assume that τ_α on $C_\alpha(Y,Z)$ satisfies* (1). *Then the following assertions are equivalent:*

(i) $C_{\beta\otimes\alpha}(T\times Y, Z) = C(T\times Y, Z)$.
(ii) $C_\alpha(Y,Z) = C(Y,Z)$ *and* $C_\beta(T, C(Y,Z)) = C(T, C(Y,Z))$ *and τ_α satisfies the exponential law for (T,Y,Z).*

PROOF. It is easy to see that $C_\alpha(Y,Z) = C(Y,Z)$. Now let $\hat{f}: T \to \to C(Y,Z)$ be continuous on each set $B \in \beta$. Let $(x_i)_i$ be a net in B converging to $x \in B$ and $(y_j)_j$ a net in $A \in \alpha$ converging to $y \in A$. The continuity of $\hat{f}$ on B means that $\hat{f}(x_i)$ converges to $\hat{f}(x)$. By (1) we infer that $\hat{f}(x_i)(y_j) := f(x_i, y_j)$ converges to $\hat{f}(x)(y) = f(x,y)$. Thus $f \in C_{\beta\otimes\alpha}(T\times Y, Z)$ and (i) yields the continuity of f. Thus τ_α satisfies the exponential law for T. Now let $(x_i)_i$ be an arbitrary net in T converging to $x \in T$. Then the continuity of $f: T\times Y \to Z$ shows that $\hat{f}(x_i)$ converges α-continuously to $\hat{f}(x)$. Proposition 1 yields the continuity of $\hat{f}: T \to C(Y,Z)$.

For the converse let $f \in C_{\beta\otimes\alpha}(T\times Y, Z)$. Since $T = \cup_{B\in\beta} B$ we have $\hat{f}(x) \in \in C_\alpha(Y,Z)$. By (ii) it is enough to show that $\hat{f} \in C_\beta(T, C_\alpha(Y,Z))$. Let $(x_i)_i$ be a net in $B \in \beta$ converging to $x \in B$. We have to show that $\hat{f}(x_i)$ converges

to $\hat{f}(x)$. By Proposition 1 this is the case if $\hat{f}(x_i)$ converges α-continuously to $\hat{f}(x)$, i.e. that $\hat{f}(x_i)(y_j) = f(x_i, y_j)$ converges to $\hat{f}(x)(y) = f(x, y)$ for every net $(y_j)_j$ in $A \in \alpha$. But this is just the continuity of $f\colon B \times A \to Z$.

Let Y be a basic locally compact space and T be an l-space. Since Condition (ii) is fulfilled for $\beta = l$ and $\alpha = k$ for all spaces Z we conclude that the product $T \times Y$ is an l-space since $C_{l\otimes l}(T \times Y, Z) \subset C_{l\otimes k}(T \times Y, Z)$. Similarly we obtain the following improvement of Theorem 3.3 in [8]: the product $T \times Y$ of a k_3-space T and a locally compact space Y is a k_3-space.

Now let Y be a regular l-space (or k-space) which is not locally compact. Endow $Z := \{0, 1\}$ with the Sierpiński topology $\{\emptyset, \{0\}, Z\}$. Then $T := (C(Y, Z), \tau_k)$ is compact, cf. [1]. Suppose that $T \times Y$ is an l-space. Then we have $C_{l\otimes k}(T \times Y, Z) = C(T \times Y, Z)$ since every bounded subset of Y is relatively compact. By Theorem 2 τ_k satisfies the exponential law for (T, Y, Z). But the identity function $\mathrm{id}\colon T \to C(Y, Z)$ is continuous and $\hat{e} = \mathrm{id}$, i.e. the evaluation $e\colon T \times Y \to Z$ is continuous. Then Y is locally compact (cf. [5]), a contradiction.

COROLLARY 3. *Let Y be regular. Then Y is locally compact if and only if $X \times Y$ is an l-space (or k-space) for all compact spaces X.*

REFERENCES

[1] FELL, J. M., A Hausdorff topology for the closed subsets of a locally compact non-Hausdorff space, *Proc. Amer. Math. Soc.* **13** (1962), 472–476. *MR* **25** #2573

[2] GUTHRIE, J. A., A characterization of $\aleph_0$-spaces, *General Topology and Appl.* **1** (1971), 105–110. *MR* **44** #5922

[3] LAMBRINOS, P., Boundedly generated topological spaces, *Manuscripta Math.* **31** (1980), 425–438. *MR* **83** m:54024

[4] LAMBRINOS, P., The bounded-open topology on function spaces, *Manuscripta Math.* **36** (1981/82), 47–66. *MR* **83m**:54025

[5] LAMBRINOS, P., On the exponential law for function spaces equipped with the compact-open topology, *Continuous lattices and their applications* (Bremen, 1982), Lecture Notes in Pure and Appl. Math. **101**, Dekker, New York, 1985, 181–190. *MR* **88a**:54038

[6] MCCOY, R. A. and NTANTU, I., *Topological properties of spaces of continuous functions*, Lecture Notes in Mathematics, **1315**, Springer, Berlin–Heidelberg–New York, 1988. *MR* **90a**:54046

[7] MICHAEL, E., Local compactness and Cartesian products of quotient maps and k-spaces, *Ann. Inst. Fourier (Grenoble)* **18** (1968), 281–286. *MR* **39** #6256

[8] MORALES, P., Non-Hausdorff Ascoli theory, *Dissertationes Math. (Rozprawy Mat.)* **119** (1974), 1–37. *MR* **53** #3996

[9] VOGT, R. M., Convenient categories of topological spaces for homotopy theory, *Arch. Math. (Basel)* **22** (1971), 545–555. *MR* **45** #9323

(Received May 27, 1990)

FACHBEREICH 11, MATHEMATIK
UNIVERSITÄT-GH DUISBURG
LOTHARSTRASSE 65
POSTFACH 10 15 03
D–47048 DUISBURG 1
FEDERAL REPUBLIC OF GERMANY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 457–463

A NOTE ON THE ALGEBRAIC DERIVATIVE AND INTEGRAL IN A DISCRETE OPERATIONAL CALCULUS

T. FÉNYES

In the paper [1] we gave a discrete operational calculus based on the number-theoretical Dirichlet product of functions defined on the positive integers. We introduced a Mikusiński-type operator field as follows. Let Z, R, K, E denote the set of the natural numbers, positive rational numbers, the complex numbers, and the ring of the real-valued functions defined in Z, respectively. The ring operations were introduced by the usual addition and the Dirichlet product by

$$ab = \Big\{\sum_{\nu|n} a(\nu)b\Big(\frac{n}{\nu}\Big)\Big\}, \qquad a, b \in E,\ n = 1, 2, \dots, \tag{1}$$

where M denotes the field of the Mikusiński operators based on the product (1).

The operator function $\delta(\alpha)$, $\alpha \in R$ was defined by $\delta(\alpha) = \frac{\delta(N_1)}{\delta(N_2)} \in M$, $\alpha = \frac{N_1}{N_2}$, $N_1, N_2 \in Z$, where

$$\delta(N) = \{\delta(n, N)\}, \quad N \in Z,$$

$$\delta(n, N) = \begin{cases} 0, & \text{for } n \neq N \\ 1, & \text{for } n = N. \end{cases}$$

$K \subset E \subset M$, and the common unit element of K, E, M is $\delta(1) = 1$. We denoted by E^* the subring of M whose elements are of the form

$$x = \frac{a}{\delta(\alpha)}, \qquad \alpha \in R, a \in E. \tag{2}$$

Obviously, $E \subset E^*$.

We defined the exponential function e^f, for $f \in E$ by its operational Taylor series (which converges pointwise for every f).

1991 *Mathematics Subject Classification.* Primary 44A40; Secondary 11A99.
Key words and phrases. Operational calculus, number theory.

Research partially supported by the Hungarian National Foundation for Scientific Research Grant No. 6032/6319.

Akadémiai Kiadó, Budapest

The definition of the algebraic derivative is

$$D(a) = \{-\log n \cdot a(n)\}, \qquad a \in E$$

$$D(x) = \frac{bD(a) - aD(b)}{b^2}, \qquad a, b \in E,\ b \neq 0,\ x = \frac{a}{b} \tag{3}$$

(see also [2]).

If $\alpha \in K$, then $D(\alpha) = 0$. Moreover,

$$D[\delta(\beta)] = -\log \beta\, \delta(\beta), \qquad \beta \in R,$$

$$D\left[\frac{a}{\delta(\varepsilon)}\right] = \frac{\left\{-\log \frac{n}{\varepsilon} \cdot a(n)\right\}}{\delta(\varepsilon)} \in E^*, \quad a \in E,\ \varepsilon \in R$$

and

$$D(e^f) = D(f)e^f, \quad f \in E.$$

The algebraic integral denoted by $\int$ is the inverse of D.

In the paper [1] we proved the following *statements.*

I. *If $x \in E^*$ and $D(x) = 0$, then x is an arbitrary number.*

II. *Let $a \in E$, $\varepsilon \in R$. $\int \frac{a}{\delta(\varepsilon)}$ exists in the ring E^* if and only if $\varepsilon \notin Z$, or $\varepsilon \in Z$ and $a(\varepsilon) = 0$. Moreover,*

$$\int \frac{a}{\delta(\varepsilon)} = \frac{\left\{-\frac{a(n)}{\log n/\varepsilon}\right\}}{\delta(\varepsilon)} + c, \qquad c \in K, \tag{4}$$

where in the case of $\varepsilon \in Z$ the symbol $\frac{a(\varepsilon)}{\log \frac{\varepsilon}{\varepsilon}}$ denotes an arbitrary number. Consequently, if $\gamma \in K$, $\gamma \neq 0$, then $\int \gamma$ does not exist in E^*.

III. *The algebraic differential equation*

$$D(x) - fx = 0, \qquad f \in E, \tag{5}$$

has a nontrivial solution in E^* if and only if $e^{-f(1)} \in R$, moreover, the general solution of (5) has the form

$$x = c\delta(e^{-f(1)}) \exp\left[\int (f - f(1))\right]. \tag{6}$$

We have stated in [3], [4] that the above statements hold in the whole discrete operator field M, too. Moreover, we have used these statements in this generalized version. In this note we will show that this generalization can be made. The following holds:

THEOREM. *In Statements I, II, III, E^* can be replaced by M.*

PROOF. I. We follow Mikusiński [5]. Let $x = \frac{p}{q}$ $(p, q \in E)$, and let $D\left(\frac{p}{q}\right) = 0$. We have

$$D(p)q - pD(q) = 0 \tag{7}$$

and

$$D^2(p)q - pD^2(q) = 0. \tag{8}$$

From (7), (8) we have

$$D^2(p)D(q) - D(p)D^2(q) = 0. \tag{9}$$

By differentiating it follows

$$D^3(p)D(q) - D(p)D^3(q) = 0. \tag{10}$$

From (7), (10) we obtain

$$D^3(p)q - D^3(q)p = 0. \tag{11}$$

Generally, we have

$$D^m(p)q - pD^m(q) = 0, \qquad m = 1, 2, \ldots . \tag{12}$$

By the definition of the algebraic derivative we can write

$$\sum_{\nu|n} (-\log \nu)^m \left[p(\nu)q\left(\frac{n}{\nu}\right) - p\left(\frac{n}{\nu}\right)q(\nu)\right] = 0, \tag{13}$$

$$n = 1, 2, \ldots , \quad m = 1, 2, \ldots .$$

Let

$$F\left(\nu, \frac{n}{\nu}\right) = p(\nu)q\left(\frac{n}{\nu}\right) - p\left(\frac{n}{\nu}\right)q(\nu).$$

We show that for every fixed n and $\nu \mid n$ $F(\nu, \frac{n}{\nu}) = 0$. If n is a prime number we choose $m = 1$ and obtain

$$-\log n\, F(n, 1) = 0.$$

Let us fix n and let $r_1, r_2, \ldots , r_k$ be the divisors of n $(r_1 > 1, r_k = n)$. From (13) we have

$$\begin{aligned} &\sum_{i=1}^{k} \log r_i F\left(r_i, \frac{n}{r_i}\right) = 0, \\ &\sum_{i=1}^{k} (\log r_i)^2 F\left(r_i, \frac{n}{r_i}\right) = 0, \\ &\qquad \vdots \\ &\sum_{i=1}^{k} (\log r_i)^k F\left(r_i, \frac{n}{r_i}\right) = 0. \end{aligned} \tag{14}$$

(14) can be considered as a homogeneous system of algebraic equations for the unknowns $F(r_i, \frac{n}{r_i})$, $i = 1, 2, \ldots, k$. Since the determinant of (14)

$$\begin{vmatrix} \log r_1 & \log r_2 \ldots \log r_k \\ (\log r_1)^2 & (\log r_2)^2 \ldots (\log r_k)^2 \\ \vdots & \\ (\log r_1)^k & (\log r_2)^k \ldots (\log r_k)^k \end{vmatrix} \neq 0,$$

we obtain that

$$p(\nu)q\left(\frac{n}{\nu}\right) - p\left(\frac{n}{\nu}\right)q(\nu) = 0, \tag{15}$$

or

$$p(\nu)q(\sigma) - p(\sigma)q(\nu) = 0, \qquad \nu \geqq 1,\ \sigma \geqq 1.$$

Since $q \neq 0$, there exists a value of σ such that $q(\sigma) \neq 0$. So

$$p(\nu) = \alpha q(\nu) \quad \text{for} \quad 1 \leqq \nu < \infty, \tag{16}$$

where $\alpha = \frac{p(\sigma)}{q(\sigma)}$.

COROLLARY. *Let $x_1, x_2, y \in M$ and let*

$$\int y = x_1, \qquad \int y = x_2,$$

then

$$x_1 = x_2 + c, \qquad c \in K.$$

II. It follows from the Corollary of I that we have to show only that $\int \frac{a}{\delta(\varepsilon)}$ does not exist in M if $a(\varepsilon) \neq 0$ $(\varepsilon \in Z)$. We have

$$\begin{aligned} \int \frac{a}{\delta(\varepsilon)} &= \int \frac{\{a(n) - a(\varepsilon)\delta(\varepsilon) + a(\varepsilon)\delta(\varepsilon)\}}{\delta(\varepsilon)} = \\ &= \int \frac{\{a(n) - a(N)\delta(N)\}}{\delta(\varepsilon)} + \int a(\varepsilon). \end{aligned} \tag{17}$$

From Statement II we see that the first integral on the right-hand side of (17) exists, so we must prove that $\int a(\varepsilon)$ does not exist.

Let γ be an arbitrary number $(\gamma \neq 0)$ and let us assume that $\exists x = \frac{a}{b} \in M$, $a, b \in E$ such that

$$D(x) = \gamma. \tag{18}$$

If $b(1) \neq 0$, then $x \in E$ and we have

$$-\log n\, x(n) = \gamma \quad \text{for} \quad n = 1, 2, \ldots,$$

and $\gamma = 0$. This is a contradiction. So $x \notin E$. Let $b(1) = b(2) = b(3) = \ldots = b(N-1) = 0$, $b(N) \neq 0$. From (18) we obtain

$$\sum_{\nu|n} \log \frac{n}{\nu^2} a(\nu) b\left(\frac{n}{\nu}\right) = \gamma \sum_{\nu|n} b(\nu) b\left(\frac{n}{\nu}\right), \qquad n = 1, 2, \ldots. \tag{19}$$

For $n = N$ we have $\log N \cdot a(1) b(N) = 0$ and $a(1) = 0$. For $n = 2N$ we have $\log \frac{N}{2} a(2) b(N) = 0$ and $a(2) = 0$.

If we continue this procedure for $n = 3N, 4N, \ldots$, so for $n = (N-1)N$ we have

$$\log \frac{N}{N-1} a(N-1) b(N) = 0 \quad \text{and} \quad a(N-1) = 0.$$

For $n = N^2$ we have $0 = \gamma b^2(N)$, a contradiction. So $\int \gamma$ does not exist.

III. Let us consider the algebraic differential equation

$$D(x) - fx = 0, \qquad f \in E. \tag{20}$$

Let $e^{-f(1)} \in R$. Then

$$x = \delta(e^{-f(1)}) \exp\left[\int (f - f(1))\right]$$

is a solution of (20). Let $y \neq 0$ be an arbitrary solution of (20). So

$$D(y) - fy = 0. \tag{21}$$

We have

$$\frac{D(x)}{D(y)} = \frac{x}{y},$$

and

$$yD(x) - xD(y) = 0.$$

We get

$$D\left(\frac{x}{y}\right) = \frac{yD(x) - xD(y)}{y^2} = 0.$$

The Corollary of I gives that $x = Cy$, $C \in K$, and the general solution of (20) is of the form

$$x = C\delta(e^{-f(1)}) \exp\left[\int (f - f(1))\right].$$

(In [1] we have shown that for $C \neq 0$, $x \in E$ if and only if $e^{-f(1)} \in Z$.)

Let $e^{-f(1)} \notin R$, we show that (20) has only the trivial solution. By applying the substitution

$$x = z \exp\left[\int (f - f(1))\right], \qquad z \in M$$

we get

$$D(z) - f(1)z = 0. \tag{22}$$

Let us assume that $z = \frac{a}{b}$ $(a, b \in E,\ b \neq 0)$ is a solution of (22). If $b(1) \neq 0$, then $z \in E$ and we have

$$(\log n + f(1))z(n) = 0, \qquad n = 1, 2, \ldots .$$

Since $e^{-f(1)} \notin R$, $\log n + f(1)$ vanishes for no value of n, consequently, $z(n)$ and $x(n)$ vanish identically. Let $b(1) = b(2) = \ldots = b(N-1) = 0$ and $b(N) \neq 0$. From (22) it is easy to deduce the identity

$$\sum_{\nu|n} \left(\log \frac{n}{\nu^2} - f(1)\right) a(\nu) b\left(\frac{n}{\nu}\right) = 0, \qquad n = 1, 2, \ldots . \tag{23}$$

We have for $n = N$

$$(\log N - f(1))a(1)b(N) = 0,$$

so $a(1) = 0$. For $n = 2N$ we have

$$\left(\log \frac{N}{2} - f(1)\right) a(2) b(N) = 0,$$

so $a(2) = 0$. By continuing this procedure we obtain that for $n = kN$ $(k \in Z)$

$$\left(\log \frac{N}{k} - f(1)\right) a(k) b(N) = 0$$

holds. So $a(k) = 0$ and a vanishes for every value of n. Consequently, $z = x = 0$ and the Theorem is proved.

REFERENCES

[1] FÉNYES, T. and SZILÁRD, K., Über diskrete Mikusińskische Operatoren, die auf Grund der Dirichletschen Produktenformel erzeugt werden, *Studia Sci. Math. Hungar.* **11** (1976), 181–199. *MR* **81b**:44014b

[2] GESZTELYI, E., The application of the operational calculus in the theory of numbers, *Number Theory* (Colloq., János Bolyai Math. Soc., Debrecen, 1968), Colloq. Math. Soc. J. Bolyai **2**, North-Holland, Amsterdam, 1970, 51–104. *MR* **42** #5922

[3] FÉNYES, T., On a discrete nonlinear operational differential equation system based on the Dirichlet product, *Studia Sci. Math. Hungar.* **22** (1987), 471–484. *MR* **89g**:44006

[4] FÉNYES, T., On an operational differential equation system, *Studia Sci. Math. Hungar.* **24** (1989), 365–375. *MR* 92i:44002

[5] MIKUSIŃSKI, J. and BOEHME, K., *Operational calculus*, Vol. II, Second edition, International Series of Monographs in Pure and Applied Mathematics, **110**, Pergamon Press, Oxford–Elmsford, N. Y.; PWN – Polish Scientific Publishers, Warsaw, 1987. *MR* 88k:44010

(Received July 10, 1990)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 465–471

A NOTE ON A NONLINEAR OPERATIONAL DIFFERENTIAL EQUATION SYSTEM

T. FÉNYES

In this paper we consider the operational differential equation system

$$\begin{aligned} D(x) &= -x(x+y)f, \\ D(y) &= y(x+y)f \end{aligned} \tag{1}$$

in the operator field M_D of Mikusiński-type based on the Dirichlet product of functions defined on the set of the natural numbers. Here D denotes the algebraic derivative, and f is an arbitrarily given real-valued function.

We prove existence criteria for the operational and function solutions of (1), and give the explicit form of these solutions. Though we are interested only in real (operational) solutions of the system (1), we can find these solutions by introducing the complex operators, i.e. the discrete operators that are based on the complex valued functions defined on the above set.

The following notations are used: Z, R, K, E, E_c denote the set of natural numbers, positive rational numbers, the complex numbers, the ring of the real-valued functions defined in [2], the ring of complex valued functions defined on the natural numbers $n = 1, 2, 3, \ldots$, respectively.

The algebraic derivative is defined as follows:

$$D(a) = \{-\log n\, a(n)\}, \quad a \in E_c,\ n = 1, 2, \ldots,$$
$$D(x) = \frac{bD(a) - aD(b)}{b^2}, \quad a, b \in E_c,\ x = \frac{a}{b}.$$

An operator x is real if $x = \frac{a}{b}$, $a, b \in E$.

The reader can find the elements of the discrete operational calculus, and the properties of the algebraic derivative and integral in the papers [1], [2], [3], [4]. However, for the sake of easy reading we give here the following theorem referring to the first order homogeneous algebraic differential equations.

THEOREM. *The algebraic differential equation*

$$D(w) - gw = 0, \qquad g \in E_c \tag{2}$$

1991 *Mathematics Subject Classification.* Primary 44A40.
Key words and phrases. Operational calculus.

Research partially supported by the Hungarian National Foundation for Scientific Research Grant No. 6032/6319.

has a nontrivial solution in M_D if and only if $g(1)$ is real and $e^{-g(1)} \in R$. The general solution of (2) *is of the form*

$$w = \gamma\,\delta(e^{-g(1)})\exp\left[\int (g - g(1))\right], \quad \gamma \in K, \tag{3}$$

moreover, for $\gamma \neq 0$, $w \in E_c$ if and only if $e^{-g(1)} \in Z$. In (3) *the integral denotes the algebraic integral (which is the inverse of D), and we agree that $\int(f - f(1))$ is that integral of $f - f(1)$ which has the value zero for $n = 1$, the exponential function is defined by its pointwise convergent operational Taylor series, so the exponential function occurring in* (3) *takes the value one for $n = 1$, $\delta(e^{-g(1)})$ is defined by the quotient of "discrete" Dirac functions:*

$$\delta(e^{-g(1)}) = \frac{\delta(M)}{\delta(N)} \quad \text{for} \quad e^{-g(1)} = \frac{M}{N}, \quad M, N \in Z,$$
$$\delta(M) = 1 \quad \text{for} \quad n = M, \text{ and zero for } n \neq M,$$
$$\delta(N) = 1 \quad \text{for} \quad n = N, \text{ and zero for } n \neq N.$$

$\delta(1) = 1$ is the unit element of the field M_D.

We say that (1) has a solution in M_D, E_c, E, respectively, if $\exists x, y \in M_D$, E_c, E, respectively, satisfy (1).

From (1) we have

$$yD(x) = -xy(x + y)f,$$
$$xD(y) = xy(x + y)f,$$

and

$$xD(y) + yD(x) = D(xy) = 0,$$

so by an elementary rule of the operational calculus

$$xy = -c, \quad c \in K$$

(see [6]). If $c = 0$, then $x = 0$ or $y = 0$.

$$\begin{aligned} &\text{For } x = 0 \qquad D(y) = y^2 f, \\ &\text{and for } y = 0 \qquad D(x) = -x^2 f. \end{aligned} \tag{4}$$

The equations (4) are algebraic Bernoulli equations. A detailed discussion of the Bernoulli equation can be found in the paper [4], therefore in our paper we assume $c \neq 0$. However, since we look for the real solutions of (1), we also assume that c is real.

Substituting $xy = -c$ to the first equation of (1) we obtain the Riccati equation

$$D(x) + fx^2 = cf \tag{5}$$

having as particular solution $x = \sqrt{c} \in K$. So $x = \pm\sqrt{c}$, $y = \mp\sqrt{c}$ are function solutions of the system (1). We call these the trivial solutions of (1), which are real if $c > 0$. Let us determine the nontrivial solutions of (1). By applying the substitution

$$x = \sqrt{c} + \frac{1}{z}, \qquad z \in M_D, \tag{5'}$$

(5) can be reduced to the linear inhomogeneous algebraic differential equation

$$D(z) - 2\sqrt{c} f z = f \tag{6}$$

having the particular solution $z = -\frac{1}{2\sqrt{c}}$.

Here we distinguish the cases $f(1) = 0$, $f(1) \neq 0$.

I. $f(1) = 0$. By applying the Theorem we have that the general solution of (6) is of the form

$$z = -\frac{1}{2\sqrt{c}} + \beta e^{2\sqrt{c}\int f}, \qquad \beta \in K,$$

consequently

$$x = \sqrt{c} + \frac{1}{\beta e^{2\sqrt{c}\int f} - \frac{1}{2\sqrt{c}}}.$$

After some calculation we obtain with $\varrho = 2\sqrt{c}\beta$

$$x = \sqrt{c}\frac{\varrho e^{2\sqrt{c}\int f} + 1}{\varrho e^{2\sqrt{c}\int f} - 1}, \tag{7}$$

and

$$y = -\sqrt{c}\frac{\varrho e^{2\sqrt{c}\int f} - 1}{\varrho e^{2\sqrt{c}\int f} + 1}. \tag{8}$$

First let $c > 0$. (7), (8) are real operators if we choose ϱ to be real. Excluding the value $\varrho = 0$ (7), (8) give the general real nontrivial operational solution of (1) in M_D. Since the functions $\varrho e^{2\sqrt{c}\int f} + 1$, $\varrho e^{2\sqrt{c}\int f} - 1$ take the value $\varrho + 1$ and $\varrho - 1$, respectively, for $n = 1$, it is easily seen by an elementary property of the Dirichlet product that $x \in E$ iff $\varrho \neq 1$, $y \in E$ iff $\varrho \neq -1$. Consequently, the solution (7), (8) is the general, nontrivial function solution of (1) if and only if $|\varrho| \neq 1$. Let now $c < 0$. In order to find the real (operational) solutions of (1) we must choose ϱ in the complex form

$$\varrho = \varrho_1 + i\varrho_2.$$

From (7) we have

$$x = i\sqrt{|c|}\frac{(\varrho_1 + i\varrho_2)\left[\cos 2\sqrt{|c|}\iint f + i\sin 2\sqrt{|c|}\iint f\right] + 1}{(\varrho_1 + i\varrho_2)\left[\cos 2\sqrt{|c|}\iint f + i\sin 2\sqrt{|c|}\iint f\right] - 1}. \tag{9}$$

It can easily be shown that the imaginary part of (9) vanishes if and only if

$$\varrho_1^2 + \varrho_2^2 = 1 \tag{10}$$

and the real solution of (1) is of the form

$$x = \sqrt{|c|}\frac{\varrho_2 \cos 2\sqrt{|c|}\iint f + \varrho_1 \sin 2\sqrt{|c|}\iint f}{1 + \varrho_2 \sin 2\sqrt{|c|}\iint f - \varrho_1 \cos 2\sqrt{|c|}\iint f}, \tag{11}$$

$$\varrho_1^2 + \varrho_2^2 = 1$$

$$y = -\sqrt{|c|}\frac{1 + \varrho_2 \sin 2\sqrt{|c|}\iint f - \varrho_1 \cos 2\sqrt{|c|}\iint f}{\varrho_2 \cos 2\sqrt{|c|}\iint f + \varrho_1 \sin 2\sqrt{|c|}\iint f}. \tag{12}$$

We show that $x \notin E$ if and only if $\varrho_1 = 1$ (then $\varrho_2 = 0$) and $y \notin E$ if and only if $\varrho_1 = -1$ ($\varrho_2 = 0$). Obviously, the function

$$1 + \varrho_2 \sin 2\sqrt{|c|}\int f - \varrho_1 \cos 2\sqrt{|c|}\int f$$

takes the value $1 - \varrho_1$ for $n = 1$. So for $\varrho_1 \neq 1$, $x \in E$ holds. Let $\varrho_1 = 1$, then $\varrho_2 = 0$ and we have

$$x = \sqrt{|c|}\frac{\sin 2\sqrt{|c|}\iint f}{1 - \cos 2\sqrt{|c|}\iint f}.$$

A Taylor series expansion gives that

$$\begin{aligned} x &= \sqrt{|c|}\frac{2\sqrt{|c|}\iint f - \frac{(2\sqrt{|c|}\int f)^3}{3!} + \frac{(2\sqrt{|c|}\int f)^5}{5!} - \dots}{1 - \left(1 - \frac{(2\sqrt{|c|}\int f)^2}{2!} + \frac{(2\sqrt{|c|}\int f)^4}{4!} - \dots\right)} \\ &= \sqrt{|c|}\frac{1 - \frac{(2\sqrt{|c|}\int f)^2}{3!} + \frac{(2\sqrt{|c|}\int f)^4}{5!} - \dots}{\frac{2\sqrt{|c|}\int f}{2!} - \frac{(2\sqrt{|c|}\int f)^3}{4!} + \dots}, \end{aligned} \tag{13}$$

so the denominator of (13) vanishes for $n = 1$, the numerator of (13) takes the value $\sqrt{|c|}$ for $n = 1$. Consequently, $x \notin E$. The function

$$\varrho_2 \cos 2\sqrt{|c|}\int f + \varrho_1 \sin 2\sqrt{|c|}\int f$$

takes the value ϱ_2 for $n=1$. So for $\varrho_2 \neq 0$, $y \in E$ holds. Let $\varrho_2 = 0$, then $\varrho_1 = 1$, or -1. If $\varrho_1 = 1$ then by (13) and $y = \frac{-c}{x}$ it follows that $y \in E$. If $\varrho_1 = -1$ then we have

$$y = \sqrt{|c|}\frac{1+\cos 2\sqrt{|c|}\,\iint f}{\sin 2\sqrt{|c|}\,\iint f}. \tag{14}$$

The denominator of (14) vanishes for $n=1$, its numerator, however, does not vanish for $n=1$. Therefore $y \notin E$. So (11), (12) is the nontrivial function solution of (1) if and only if $|\varrho_1| \neq 1$.

II. $f(1) \neq 0$. If $c<0$, or $c>0$ and $e^{-2\sqrt{c}f(1)} \notin R$, then applying the Theorem it can easily be seen that (1) has only trivial solutions.

Let $c>0$ and $e^{-2\sqrt{c}f(1)} = \frac{M}{N}$, M, N are relative primes. Applying (5'), (6) and the Theorem we obtain

$$x = \sqrt{c} + \frac{1}{\beta\,\delta(\frac{M}{N})\exp\left[2\sqrt{c}\int(f-f(1))\right] - \frac{1}{2\sqrt{c}}}, \quad \beta \in K.$$

A simple calculation gives with $\varrho = 2\sqrt{c}\beta$

$$\begin{aligned} x &= \sqrt{c}\frac{\varrho\,\delta(M)\exp\left[2\sqrt{c}\int(f-f(1))\right] + \delta(N)}{\varrho\,\delta(M)\exp\left[2\sqrt{c}\int(f-f(1))\right] - \delta(N)}, \\ y &= -\sqrt{c}\frac{\varrho\,\delta(M)\exp\left[2\sqrt{c}\int(f-f(1))\right] - \delta(N)}{\varrho\,\delta(M)\exp\left[2\sqrt{c}\int(f-f(1))\right] + \delta(N)}. \end{aligned} \tag{15}$$

Choosing ϱ to be real and $\varrho \neq 0$ (15) determines the general nontrivial real solution of (1). We show that $x, y \in E$ if and only if $M=1$, or $N=1$. If $M=1$, then $N>1$. The functions

$$\varrho\exp\left[2\sqrt{c}\int(f-f(1))\right] \pm \delta(N)$$

have the value ϱ for $n=1$. So by an elementary property of the Dirichlet product $x, y \in E$. If $N=1$, then $M>1$. Since the function $\delta(M)\exp[\ldots]$ vanishes for $n=1$, we obtain that for $n=1$ the functions

$$\varrho\,\delta(M)\exp\left[2\sqrt{c}\int(f-f(1))\right] - 1,$$

$$\varrho\,\delta(M)\exp\left[2\sqrt{c}\int(f-f(1))\right] + 1$$

have the value -1 and 1, respectively. Consequently, we have $x, y \in E$ again. Let $M \neq 1$, $N \neq 1$. Let us assume that $x = \{x(n)\} \in E$. Then by (15)

$$\begin{aligned} &\varrho x\, \delta(M) \exp\left[2\sqrt{c} \int (f - f(1))\right] - \delta(N)x = \\ &= \sqrt{c}\varrho\, \delta(M) \exp\left[2\sqrt{c} \int (f - f(1))\right] + \sqrt{c}\, \delta(N) \end{aligned} \tag{16}$$

holds. By an elementary property of the δ-functions, for every $h \in E$, and every $r \in Z$

$$\delta(r)h = \left\{h\left(\frac{n}{r}\right)\right\},$$

where $h(\frac{n}{r})$ is zero for such values of n, for which r is not a divisor of n. By introducing

$$u = \exp[\ldots], \qquad v = x \exp[\ldots] = xu$$

(16) can be written in the form

$$\varrho\, \delta(M)v - \delta(N)x = \sqrt{c}\varrho\, \delta(M)u + \sqrt{c}\, \delta(N). \tag{17}$$

Consequently,

$$\varrho v\left(\frac{n}{M}\right) - x\left(\frac{n}{N}\right) = \sqrt{c}\varrho u\left(\frac{n}{M}\right) + \sqrt{c}\, \delta(N) \tag{18}$$

holds for $n = 1, 2, \ldots$. Substituting $n = M$ we have

$$\varrho v_1 = \sqrt{c}\varrho u_1$$

and substituting $n = N$ in (18) we get

$$-x(1) = \sqrt{c}.$$

Since $u(1) = 1$, $v(1) = x(1)$, we obtain

$$x(1) = \sqrt{c}.$$

This is a contradiction, so $x \notin E$. Analogously we have that $y \notin E$.

The results may be summarized in the following

THEOREM 1. *Let us consider the nonlinear algebraic differential equation system*

$$\begin{aligned} D(x) &= -x(x + y)f \\ D(y) &= y(x + y)f \end{aligned} \tag{19}$$

in the discrete operator field M_D, where f is a given real-valued function. For every real solution x, y of (19) $xy = -c$ holds, where c is an arbitrary real number.

Let $f(1) = 0$. *For* $c > 0$, *the general real nontrivial solution of* (19) *is of the form* (7), (8) *where* $\varrho \neq 0$ *is an arbitrary real number.* (7), (8) *is the general nontrivial function solution if and only if* $|\varrho| \neq 1$. *For* $c < 0$ *the general real nontrivial solution is of the form* (11), (12), *where* ϱ_1, ϱ_2 *are arbitrary real numbers satisfying* $\varrho_1^2 + \varrho_2^2 = 1$. (11), (12) *is the general function solution of* (19) *if and only if* $|\varrho_1| \neq 1$.

Let $f(1) \neq 0$. *For* $c < 0$, *or* $c > 0$ *and* $e^{-2\sqrt{c}f(1)} \notin R$, (19) *has only trivial solutions. For* $c > 0$ *and* $e^{-2\sqrt{c}f(1)} = \frac{M}{N}$, $M, N \in Z$, *(*M, N *are relative primes), the general real nontrivial solution of* (19) *is given by* (15), *where* $\varrho \neq 0$ *is an arbitrary real number.* (15) *is the general real nontrivial function solution of* (19) *if and only if* $M = 1$, *or* $N = 1$.

REFERENCES

[1] FÉNYES, T. and SZILÁRD, K., Über diskrete Mikusińskische Operatoren, die auf Grund der Dirichletschen Produktenformel erzeugt werden, *Studia Sci. Math. Hungar.* **11** (1976), 181–199. *MR* **81b**:44014b

[2] GESZTELYI, E., The application of the operational calculus in the theory of numbers, *Number Theory* (Colloq., János Bolyai Math. Soc., Debrecen, 1968), Colloq. Math. Soc. J. Bolyai **2**, North-Holland, Amsterdam, 1970, 51–104. *MR* **42** #5922

[3] FÉNYES, T., On a discrete nonlinear operational differential equation system based on the Dirichlet product, *Studia Sci. Math. Hungar.* **22** (1987), 471–484. *MR* **89g**:44006

[4] FÉNYES, T., On an operational differential equation system, *Studia Sci. Math. Hungar.* **24** (1989), 365–375. *MR* **92i**:44002

[5] FÉNYES, T., On an algebraic differential equation of Bernoulli type, *Studia Sci. Math. Hungar.* **28** (1993), 115–129.

[6] FÉNYES, T., A note on the algebraic derivative and integral in a discrete operational calculus, *Studia Sci. Math. Hungar.* **28** (1993), 457–463.

(Received July 10, 1990)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY

Studia Scientiarum Mathematicarum Hungarica 28 (1993), 473–483

EXTERNAL ILLUMINATION ACCORDING TO L. FEJES TÓTH

V. SOLTAN

Illumination by points

In 1977 L. Fejes Tóth [1] introduced the following notion of illumination: a set $X \subset E^n$ is called illuminated by a set $Y \subset E^n \setminus X$ provided for every point $x \in \operatorname{bd} X$ there is a point $y \in Y$ such that $]x, y[\cap X = \emptyset$. A stronger type of the illumination (see [2]) is known: a set $V \subset E^n$ is called illuminated by a set $W \subset E^n \setminus V$ if for every point $v \in \operatorname{bd} V$ there is a point $w \in W$ such that $]v, w[\cap V = \emptyset$ and the ray $[w, v\rangle$ with the apex w passing through v intersects $\operatorname{int} V$.

Below these two types of the illumination will be called *weak* and *strong*, respectively.

There are some problems about the strong illumination of convex bodies (see for example [3]). In this paper we investigate analogous problems about weak illumination.

Let K be a convex body (a proper closed convex set with non-empty interior) in the n-dimensional linear space E^n. Denote by $p(K)$ (respectively, by $q(K)$) the least number of points in $E^n \setminus K$ which strongly (weakly) illuminate K. Put $p(K) = \infty$ ($q(K) = \infty$) if K cannot be strongly (weakly) illuminated by any finite number of points from $E^n \setminus K$. Obviously, $q(K) \leqq p(K)$.

The problem on the upper bound for $p(K)$ is open up to now. For a compact body K this problem is equivalent to the famous Hadwiger problem on the covering of K by the least number of smaller homothetic copies [4]. Therefore the inequality $p(K) \leqq 2^n$ is conjectured (see [3] for details).

The following result shows that the upper bound for $q(K)$ appears much simpler.

THEOREM 1. *For a compact convex body $K \subset E^n$ one has $2 \leqq q(K) \leqq n + 1$.*

PROOF. Let $S = \operatorname{conv}(a_1, \ldots, a_{n+1})$ be any n-dimensional simplex with vertices $a_1, \ldots, a_{n+1}$ such that $K \subset \operatorname{int} S$. Choose any point $x \in \operatorname{bd} K$ and

1980 *Mathematics Subject Classification* (1985 Revision). Primary 52A40.
Key words and phrases. Convex sets, external illumination.

Akadémiai Kiadó, Budapest

some hyperplane H supporting K at x. Denote by P the open half-space which is bounded by H and does not intersect K. The inclusion $K \subset \operatorname{int} S$ implies that at least one of the points $a_1, \dots, a_{n+1}$, say a_i, belongs to P. Then $]a_i, x[\subset P$, and $]a_i, x[\cap K = \emptyset$. Thus K is weakly illuminated by the set $\{a_1, \dots, a_{n+1}\}$, i.e., $q(K) \leqq n+1$. The inequality $q(K) \geqq 2$ is trivial. □

It is easy to verify that $q(B) = n+1$ for any euclidean ball $B \subset E^n$. Therefore the inequality $q(K) \leqq n+1$ is sharp.

For an unbounded convex body $K \subset E^n$ the values $p(K)$ and $q(K)$ can be infinite. For example, if $K \subset E^2$ is a convex figure bounded by a parabola, then $p(K) = q(K) = \infty$.

It is known (see [3, p. 251]) that an unbounded convex body $K \subset E^n$ is strongly illuminated by a finite number of points if and only if it is almost conic (K is called almost conic provided it is contained in some r-neighbourhood of its characteristic cone).

If $K \subset E^n$ is an almost conic unbounded convex body, then $p(K) \leqq 2$ for $n = 2$ and $p(K) \leqq 4$ for $n = 3$ (see [2] and [3], respectively).

CONJECTURE 1. *For an unbounded almost conic convex body $K \subset E^n$, one has $p(K) \leqq 2^{n-1}$.*

In case of weak illumination, the problem to describe the family of all convex bodies $K \subset E^n$, $n \geqq 3$ satisfying the condition $q(K) < \infty$ still remains open. The following example shows that this family is larger than the family of all almost conic bodies.

EXAMPLE 1. Let H be some $(n-1)$-dimensional paraboloid in E^n, $n \geqq 3$ and $x \in E^n \setminus \operatorname{aff} H$. The convex body $K = \overline{\operatorname{conv}}(x \cup H)$ is not almost conic. At the same time $q(K) = 2$.

It is easy to prove the following lemma.

LEMMA 1. *An unbounded convex figure $K \subset E^2$ is weakly illuminated by a finite number of points if and only if it is almost conic.*

THEOREM 2. *If for an unbounded convex body $K \subset E^n$ the value $q(K)$ is finite, then $1 \leqq q(K) \leqq n$.*

For the proof of Theorem 2 we need some auxiliary notions and results. Denote by H_x some hyperplane supporting K at a point $x \in \operatorname{bd} K$, and by Q_x that open half-space bounded by H_x which does not intersect K. If $N \subset \operatorname{bd} K$ is any open subset, then N_0 will denote the set of all regular points of K contained in N. For any regular point $x \in \operatorname{bd} K$ let e_x be the unit vector in E^n parallel to the external normal to K at x. Put $S_N = \{e_x : x \in N_0\}$.

LEMMA 2. *A set $N \subset \operatorname{bd} K$ is weakly illuminated by a point from $E^n \setminus K$ if and only if the set $R_N := \cap\{Q_x : x \in N_0\}$ is not empty.*

PROOF. Suppose that N is weakly illuminated by some point $z \in E^n \setminus K$. One has $z \in \overline{Q}_x$ for any point $x \in N_0$ (otherwise $]x, z[$ intersects $\operatorname{int} K$). Hence $z \in \cap\{\overline{Q}_x : x \in N_0\}$. Choose in $\operatorname{int} K$ some open ball V.

Since V does not intersect the set $\cap\{\overline{Q}_x : x \in N_0\}$, any ray $[v, z\rangle$ with apex $v \in V$ passing through z intersects each open half-space Q_x, $x \in N_0$. Hence R_N contains some open cone with apex z. So $R_N \neq \emptyset$.

Conversely, assume that $R_N \neq \emptyset$. Choose any point $w \in R_N$ and some ball $V \subset \operatorname{int} K$. As above, we show that R_N contains some open cone T with apex w. Since $[w, x[\subset Q_x$ for each point $x \in N_0$, the set N_0 is weakly illuminated by w. All singular points from N which are not weakly illuminated by w may be situated only on the boundary of the cone with apex w generated by K (denote this cone by W). We choose any point $y \in \operatorname{int} T \setminus W$. It is easy to see that y weakly illuminates the whole set N. □

Proof of Theorem 2. Suppose that $q(K) < \infty$, and let $z_1, \ldots, z_m$ be some points in $E^n \setminus K$ weakly illuminating K. Denote by N_i all points in $\operatorname{bd} K$ weakly illuminated by z_i, $i = 1, \ldots, m$. By Lemma 2, one has $R_{N_i} \neq \emptyset$, $i = 1, \ldots, m$.

K being unbounded, contains a ray, l. Denote by S the set of all unit vectors which are translates of the external normals to K at its regular points. Clearly, if f is the unit vector in l, then $(e, f) \leqq 0$ for any $e \in S$. Hence S belongs to some closed hemisphere Φ of the unit sphere in E^n.

It is possible to cover Φ by some n closed spherical segments $\Phi_1, \ldots, \Phi_n$ smaller than a hemisphere. Denote by l_j the ray with apex O which lies inside Φ_j in the axis of symmetry of Φ_j, and by e_j the unit vector in l_j. Put $S_{ij} = S_{N_i} \cap \Phi_j$, $i = 1, \ldots, m$, $j = 1, \ldots, n$ and

$$S_j = \cup\{S_{ij} : 1 \leqq i \leqq m\}.$$

Denote by M_j (M_{ij}) the set of all regular points $s \in \operatorname{bd} K$ for which the corresponding vector e_s belongs to S_j (S_{ij}), and put

$$R_j = \cap\{Q_x : x \in M_j\}, \quad R_{ij} = \cap\{Q_x : x \in M_{ij}\}.$$

Since $R_{N_i} \neq \emptyset$, each of the sets R_{ij} is not empty. We have $(e_j, t) > 0$ for any vector $t \in S_{ij}$. Hence the ray l_j intersects each of the sets R_{ij}, $i = 1, \ldots, m$. Therefore, the intersection

$$l_j \cap R_j = \cap\{l_j \cap R_{ij} : 1 \leqq i \leqq m\}$$

is non-empty. By Lemma 2, the closed set $\overline{M}_j$ is weakly illuminated by some point w_j. The obvious relation $\operatorname{bd} K = \cup\{\overline{M}_j : 1 \leqq j \leqq n\}$ implies that the set $\{w_1, \ldots, w_n\}$ weakly illuminates K, i.e., $q(K) \leqq n$. □

A set of points illuminating a convex body K is called *primitive* if none of its proper subsets illuminates K (see [5, p. 422]). Denote by $\overline{p}(K)$ (respectively, by $\overline{q}(K)$) the supremum of cardinalities of primitive sets strongly (weakly) illuminating K. Put $\overline{p}(K) = \infty$ (respectively, $\overline{q}(K) = \infty$) if the supremum is infinite.

Obviously, $p(K) \leqq \overline{p}(K)$ and $q(K) \leqq \overline{q}(K)$.

Observation 1. Unlike the inequality $q(K) \leqq p(K)$, the relation $\overline{q}(K) \leqq \overline{p}(K)$ is not true for any body K (compare Theorems 3 and 4).

THEOREM 3. *For a convex body $K \subset E^n$, one has $\overline{q}(K) < \infty$ if and only if K is polyhedral. If K is polyhedral, then $\overline{q}(K)$ equals the number $f^{n-1}(K)$ of all facets of K.*

PROOF. Suppose that K is not polyhedral. Then for any natural number m it is possible to find m regular points $x_1, \dots, x_m \in \operatorname{bd} K$ such that the hyperplanes $H_1, \dots, H_m$ supporting K at the points $x_1, \dots, x_m$, respectively, are different. Denote by Q_i that open half-space which is bounded by H_i and does not intersect K. For each point x_i we can choose a point $z_i \in Q_i \setminus (\cup \overline{Q}_j : j \neq i)$ weakly illuminating some open neighbourhood U_i of x_i on $\operatorname{bd} K$. Put

$$Z = \{z_1, \dots, z_m\}, \quad W = E^n \setminus (K \cup \overline{Q}_1 \cup \dots \cup \overline{Q}_m).$$

Clearly, W weakly illuminates the closed set

$$T := \operatorname{bd} K \setminus (U_1 \cup \dots \cup U_m).$$

Following the standard compactness arguments, we can choose in W some at most countable subset G (G is finite if K is compact) which is a primitive weakly illuminating set for T.

The set $Z \cup G$ weakly illuminates K, and each point $w \in G$ cannot weakly illuminate any of the points $x_1, \dots, x_m$. Therefore any subset of $Z \cup G$ weakly illuminating K contains Z. Hence any primitive weakly illuminating set contained in $Z \cup G$ has at least m elements. The last shows that $\overline{q}(K) \geqq m$. Since m is arbitrary, one has $\overline{q}(K) = \infty$.

Suppose that K is polyhedral, and put $t = f^{n-1}(K)$. Near each facet F_i of K we can place a point $y_i \in E^n \setminus K$ sufficiently close to F_i such that y_i weakly illuminates only F_i, $i = 1, \dots, t$. Then $\{y_1, \dots, y_t\}$ is a primitive weakly illuminating set for K. Hence $\overline{q}(K) \geqq f^{n-1}(K)$.

Conversely, let Y be any primitive weakly illuminating set for K. For each facet F_i of K we choose some point $x_i \in \operatorname{rint} F_i$ and a point $w_i \in Y$ weakly illuminating x_i, $i = 1, \dots, t$. Then w_i weakly illuminates the whole facet F_i. So the set $\{w_1, \dots, w_t\}$ weakly illuminates K. One has $X = \{w_1, \dots, w_t\}$, since Z is primitive. Hence $\overline{q}(K) \leqq f^{n-1}(K)$. □

COROLLARY 1. *For a convex body $K \subset E^n$, one has $\overline{q}(K) \geqq 1$. The equality $\overline{q}(K) = 1$ holds if and only if K is a half-space. If K is compact, then $\overline{q}(K) \geqq n+1$ with $\overline{q}(K) = n+1$ only for simplices.*

Now we shall study the value $\overline{p}(K)$. A suitable description for $\overline{p}(K)$ can be realized for line-free convex bodies. Therefore the general case will be reduced to this one.

It is well-known that each convex body $K \subset E^n$ can be uniquely represented (up to an affinity) as a direct sum $K = L + M$, where L is a linear subspace and M is a line-free closed convex set.

THEOREM 4. *Suppose that a convex body $K \subset E^n$ is represented as a direct sum $K = L + M$, where L is a linear subspace and M is a line-free closed convex set. Then $\overline{p}(K) < \infty$ if and only if M has a finite number of extremal points. In this case $\overline{p}(K) = \operatorname{card} \operatorname{ext} M$.*

We divide the proof of Theorem 4 into a sequence of Lemmas 3–5.

LEMMA 3. *One has $\overline{p}(K) = \overline{p}(M)$, where $\overline{p}(M)$ means the corresponding number for M in the linear space* aff M.

PROOF. $p(K) = p(M)$ if both numbers $\overline{p}(K), \overline{p}(M)$ are infinite. Suppose that $\overline{p}(K) < \infty$, and let some points $v_1, \ldots, v_k \in E^n$ form a primitive strongly illuminating set for K. Then the projections of these points on aff M parallel to L form a primitive strongly illuminating set for M in the space aff M. Hence $\overline{p}(K) \leqq \overline{p}(M)$.

Conversely, if some points $w_1, \ldots, w_m \in \operatorname{aff} M$ form a primitive strongly illuminating set for M, then, also, they form a primitive strongly illuminating set for K in E^n. Hence $\overline{p}(M) \leqq \overline{p}(K)$. □

LEMMA 4. *A set X strongly illuminates a line-free convex body $K \subset E^n$ if and only if X strongly illuminates the set* ext K.

PROOF. The well-known assertion of V. Klee [6] states that for any line-free convex body $K \subset E^n$ the following relation holds: $K = \operatorname{conv}(\operatorname{ext} K \cup \cup \operatorname{rext} K)$, where rext K means the union of all extremal rays of K. This assertion makes it obvious that any point $z \in \operatorname{bd} K \setminus \operatorname{ext} K$ belongs to some open interval $]v, w[\subset \operatorname{bd} K$ such that $v \in \operatorname{ext} K$.

Now we turn to the proof of Lemma 4. If a set X strongly illuminates K, then X strongly illuminates the set ext K. Conversely, let a set X strongly illuminate ext K, and $z \in \operatorname{bd} K$ be any point which is not extremal for K. By the above, z belongs to some open interval $]v, w[\subset \operatorname{bd} K$ such that $v \in \operatorname{ext} K$. A trivial verification shows that if some point $x \in X$ strongly illuminates v, then x strongly illuminates z. Hence X strongly illuminates K. □

LEMMA 5. *Let $K \subset E^n$ be a line-free convex body. Then $\overline{p}(K) < \infty$ if and only if K has a finite number of extremal points. In this case $\overline{p}(K) =$ $= \operatorname{card} \operatorname{ext} K$.*

PROOF. First we shall demonstrate the validity of the following assertion: if $\operatorname{card} \operatorname{ext} K \geqq m$, then $\overline{p}(K) \geqq m$.

From the assumption $\operatorname{card} \operatorname{ext} K \geqq m$ and the well-known relation ext $K \subset$ $\subset \overline{\exp} K$ there follows the inequality $\operatorname{card} \exp K \geqq m$. Let $x_1, \ldots, x_m$ be some exposed points of K, and $L_1, \ldots, L_m$ be some closed half-spaces in E^n such that $L_i \cap K = \{x_i\}$, $i = 1, \ldots, m$. For each point x_i we can choose a point $z_i \in L_i \setminus (\cup L_j : j \neq i)$ strongly illuminating some open neighbourhood U_i of x_i on bd K. Put

$$Z = \{z_1, \ldots, z_m\}, \qquad W = E^n \setminus (K \cup L_1 \cup \ldots \cup L_m).$$

Clearly, W strongly illuminates the closed set

$$T := \operatorname{bd} K \setminus (U_1 \cup \ldots \cup U_m).$$

Following the standard compactness arguments, we can choose in W some at most countable subset G (G is finite if K is compact) which is a primitive strongly illuminating set for T.

The set $Z \cup G$ strongly illuminates K, and each point $w \in G$ cannot strongly illuminate any of the points $x_1, \ldots, x_m$. Therefore any subset of $Z \cup G$ strongly illuminating K contains Z. Hence any primitive strongly illuminating set contained in $Z \cup G$ has at least m elements. The last implies that $\overline{p}(K) \geqq m$.

If K has an infinite number of extremal points, then, by the above, one has $\overline{p}(K) = \infty$.

Suppose now that K has a finite number of extremal points: $\operatorname{ext} K = \{x_1, \ldots, x_m\}$. As demonstrated above, $\overline{p}(K) \geqq m$. Let Y be any strongly illuminating set for K. We choose in Y some points $y_1, \ldots, y_m$ such that y_i strongly illuminates x_i, $i = 1, \ldots, m$. By Lemma 4, the set $\{y_1, \ldots, y_m\}$ strongly illuminates K. Hence any primitive illuminating set for K has at most m points, i.e., $\overline{p}(K) \leqq m$. □

COROLLARY 2. *For a convex body $K \subset E^n$ one has $\overline{p}(K) \geqq 1$. The equality $\overline{p}(K) = 1$ holds if and only if K is a cone. If K is compact, then $\overline{p}(K) \geqq n + 1$ and $\overline{p}(K) = n + 1$ holds only for simplices.*

Illumination by directions

V. G. Boltjanskiĭ [4] introduced the notion of illumination by (oriented) directions: a convex body $K \subset E^n$ is called illuminated by a family $\mathcal{L}$ of directions in E^n provided for each point $x \in \operatorname{bd} K$ there is a direction $l \in \mathcal{L}$ such that the ray l_x with apex x and direction l intersects $\operatorname{int} K$. This type of illumination will be called below *strong*.

Analogously to illumination in the sense of L. Fejes Tóth, we shall introduce the following type of illumination: a convex body $K \subset E^n$ is called *weakly* illuminated by a family $\mathcal{L}$ of directions in E^n provided for each point $x \in \operatorname{bd} K$ there is a direction $l \in \mathcal{L}$ such that the ray l'_x with apex x and the direction opposite to l intersects K at the point x only.

Denote by $s(K)$ (respectively, by $r(K)$) the least number of directions in E^n which strongly (weakly) illuminate K. Put $s(K) = \infty$ ($r(K) = \infty$) if K cannot be strongly (weakly) illuminated by any finite family of directions. Obviously,

$$r(K) \leqq s(K) \leqq p(K), \qquad r(K) \leqq q(K) \leqq p(K).$$

THEOREM 5. *For any convex body $K \subset E^n$, one has $r(K) \leqq 2$.*

PROOF. From [7] there follows the existence of a line $l \subset E^n$ which is not parallel to any line segment in $\operatorname{bd} K$. Note that l does not belong to K

(otherwise K would be a cylinder whose boundary contains line segments parallel to l). Denote by l_1, l_2 two opposite directions determined by l. It is easy to verify that the directions l_1, l_2 weakly illuminate K. Hence $r(K) \leqq 2$. □

The problem on the upper estimation of $s(K)$ is still open. For a compact body K it is equivalent to Hadwiger's covering problem (see [4]). Therefore the validity of the inequality $s(K) \leqq 2^n$ is conjectured. For an unbounded body K the value $s(K)$ can be infinite ($s(K) = \infty$ for K in Example 1). The description of all convex bodies $K \subset E^n$ for which $s(K) < \infty$ has not been found (this family is larger than the family of all almost conic bodies).

CONJECTURE 2. *If for an unbounded convex body $K \subset E^n$, the value $s(K)$ is finite, then $1 \leqq s(K) \leqq 2^{n-1}$.*

For any unbounded convex figure $K \subset E^2$ one has $s(K) \leqq 2$ [2].

A family of directions illuminating a convex body K is called *primitive* if none of its proper subfamilies illuminates K (see [5, p. 422]). Denote by $\overline{s}(K)$ (respectively, by $\overline{r}(K)$) the supremum of cardinalities of primitive families strongly (weakly) illuminating K. Put $\overline{s}(K) = \infty$ (respectively, $\overline{r}(K) = \infty$) if the supremum is infinite.

Obviously, $s(K) \leqq \overline{s}(K)$ and $r(K) \leqq \overline{r}(K)$.

Observation 2. Unlike the inequality $r(K) \leqq s(K)$, the relation $\overline{r}(K) \leqq \overline{s}(K)$ is not satisfied for an arbitrary convex body K.

THEOREM 6. *For a convex body $K \subset E^n$, one has $\overline{s}(K) \geqq 1$, with $\overline{s}(K) = 1$ if and only if K is a cone.*

PROOF. The first statement of the theorem is trivial. If K is a cone, then, clearly, $\overline{s}(K) = 1$.

Conversely, let $\overline{s}(K) = 1$. We need the following assertion [8]: if C is the characteristic cone of a line-free convex body N, then

$$C = \cap\{Q(N, x) - x : x \in \exp N\},$$

where $Q(N, x) := \cup\{x + \lambda(N - x) : \lambda \geqq 0\}$.

Suppose that K is not a cone. Let $K = L + M$ be a representation of K as a direct sum of a linear subspace L and a line-free closed convex set M. Then M is not a cone. By the above,

$$W := \operatorname{rint} Q(M, x) \setminus (\operatorname{rint} C(M) + x) \neq \emptyset$$

for any point $x \in \exp M$. Fix some point $a \in \exp M$, and choose in W a ray l_a with apex a. A trivial verification shows that the direction l determined by l_a strongly illuminates some open neighbourhood U of the exposed face $F := a + L$ of K, and l does not illuminate strongly the whole body K.

Let H be a hyperplane in E^n such that $H \cap K = F$, and f be a unit normal to H in the direction of K. Let $\mathcal{L}$ be the family of all directions in

E^n which form with f an angle $> \pi/2$. It is easy to verify that $\mathcal{L}$ strongly illuminates $\operatorname{bd} K \setminus U$, and no direction from $\mathcal{L}$ strongly illuminates F.

Following the standard compactness arguments, we can choose in $\mathcal{L} \cup \{l\}$ some at most countable subfamily $\mathcal{N}$ ($\mathcal{N}$ is finite if K is compact), which is a primitive strongly illuminating set for K. Hence $\overline{s}(K) \geqq \operatorname{card} \mathcal{N} \geqq 2$, which is impossible. □

THEOREM 7. *For a convex body $K \subset E^n$, one has $\overline{r}(K) \geqq 1$, with $\overline{r}(K) = 1$ if and only if K is a half-space.*

PROOF. The first statement of Theorem 7 is trivial. If K is a half-space, then, clearly, $\overline{r}(K) = 1$.

Conversely, let $\overline{r}(K) = 1$, and suppose that K is not a half-space. Then two points $a, b \in \operatorname{bd} K$ can be chosen such that $]a, b[\subset \operatorname{int} K$. Denote by l_1, l_2 the opposite directions determined by the line $\langle a, b \rangle$. We can slightly change the directions l_1, l_2 in order to illuminate the whole body K (see Theorem 5). Hence $\overline{r}(K) \geqq 2$, which is impossible. □

If a convex body $K \subset E^n$ is compact, then $s(K) \geqq n+1$ (see [3]). Therefore one has $\overline{s}(K) \geqq s(K) \geqq n+1$ if K is compact.

CONJECTURE 3. *If a convex body $K \subset E^n$ is compact, then $\overline{r}(K) \geqq n+1$.*

The problem to determine the least upper bound for $\overline{s}(K)$ is posed by B. Grünbaum [5, p. 423], who observed that $\overline{s}(K) \leqq 6$ for any convex figure $K \subset E^2$, with $\overline{s}(K) = 6$ holding only for hexagons with parallel opposite sides. It is easy to see that $\overline{s}(K) \leqq 3$ for an unbounded convex figure $K \subset E^2$.

Note that $\overline{r}(K) \leqq 4$ for any convex figure $K \subset E^2$, and $\overline{r}(K) \leqq 3$ if K is unbounded.

B. Grünbaum [5, p. 423] conjectured the validity of the inequality $\overline{s}(K) \leqq 2(2^n - 1)$ for any compact convex body $K \subset E^n$, $n \geqq 3$. Below we give an example of a compact convex body $K \subset E^3$ such that $\overline{s}(K) = \overline{r}(K) = \infty$.

EXAMPLE 2. Let $Q = \{(x, y, z) : x^2 + y^2 \leqq z^2, 0 \leqq z \leqq 1\}$ be a bounded circular cone in E^3. Denote by l_0 the direction determined by a vector $(-1, 0, \varepsilon - 1)$, $0 < \varepsilon < 1$. An easy computation shows that l_0 weakly illuminates the upper disc

$$D = \{(x, y, z) : x^2 + y^2 \leqq 1, z = 1\}$$

and a closed sector C of the lateral surface of Q bounded by the generatrices

$$V_1 = \{(\lambda p, \lambda q, \lambda) : \lambda \geqq 0\}, \qquad V_2 = \{(\lambda p, -\lambda q, \lambda) : \lambda \geqq 0\},$$

where

$$p = \left(\sqrt{1 + 2\varepsilon - \varepsilon^2} + 1 - \varepsilon\right)/2, \quad q = \left(\sqrt{1 + 2\varepsilon - \varepsilon^2} - 1 + \varepsilon\right)/2.$$

Note that the width of C becomes arbitrarily small if $\varepsilon \to 0$.

Now it is clear that for arbitrary $m \geqq 4$ it is possible to choose a suitable ε such that the directions $l_0, l_1, \ldots, l_{m-1}$ determined by the respective vectors

$$(-\cos(2\pi k/m), -\sin(2\pi k/m), \varepsilon - 1), \qquad k = 0, 1, \ldots, m-1,$$

form a primitive weakly illuminating family for Q. Hence $\overline{r}(Q) = \infty$.

The direction l_0 strongly illuminates not only the interiors of D and C. Also, it illuminates strongly the least open arc of the circle $F = \{(x, y, z) : x^2 + y^2 = 1, z = 1\}$ bounded by the points $(p, q, 1), (p, -q, 1)$. Note that l_0 does not illuminate strongly the apex $(0, 0, 0)$ of Q. The direction l' determined by the vector $(0, 0, 1)$ strongly illuminates the apex of Q, but does not illuminate strongly any of the points from F.

Hence for arbitrary $m \geqq 4$ it is possible to choose a suitable ε such that the above mentioned directions $l_0, l_1, \ldots, l_{m-1}$ together with l' form a primitive strongly illuminating family for Q. So $\overline{s}(Q) = \infty$.

Observation 3. After obtaining this example, the author learned about the similar example of K. Bezdek for the case of strong illumination. This will appear in his paper "Hadwiger–Levi's covering problem revisited", submitted to the book edited by J. Pach *Recent progress in discrete and computational geometry* at the Springer-Verlag.

External visibility

F. A. Valentine [9] introduced the notion of external visibility, which coincides with the notion of illumination according to L. Fejes Tóth. Below we study briefly a weaker type of visibility introduced by E. Buchman and F. A. Valentine [10, 11].

Let K be a convex body in E^n. We say that a point $x \in E^n \setminus K$ sees a set $N \subset \operatorname{bd} K$ (or N is visible from x) if $[x, y] \cap \operatorname{int} K = \emptyset$ for any point $y \in N$. A set $X \subset E^n \setminus K$ sees K provided for any point $y \in \operatorname{bd} K$ there is a point $x \in X$ which sees y.

Denote by $v(K)$ the least cardinality of a set in $E^n \setminus K$ which sees the whole body K. Put $v(K) = \infty$ if any finite set $X \subset E^n \setminus K$ cannot see K. Obviously, $v(K) \leqq q(K)$.

THEOREM 8. *For any convex body $K \subset E^n$ one has $v(K) = q(K)$.*

PROOF. It is sufficient to prove the inequality $q(K) \leqq v(K)$. Suppose that some set $\{z_1, \ldots, z_m\} \subset E^n \setminus K$ sees K. For any $i = 1, \ldots, m$ denote by C_i the cone with apex z_i generated by K:

$$C_i = \cup\{z_i + \lambda(K - z_i) : \lambda \geqq 0\}.$$

Let C_i' denote the cone with apex z_i symmetric to C_i: $C_i' = 2z_i - C_i$. An easy verification shows that for any points $y_i \in \operatorname{int} C_i'$, $i = 1, \ldots, m$, the set $\{y_1, \ldots, y_m\}$ weakly illuminates K. Hence $q(K) \leqq v(K)$. □

A set of points which sees a convex body K is called *primitive* if none of its proper subsets sees K. Denote by $\overline{v}(K)$ the supremum of cardinalities of primitive sets which see K. Analysis of the proof of Theorem 3 shows that it remains true if we shall substitute $\overline{q}(K)$ by $\overline{v}(K)$. Therefore $\overline{v}(K) = \overline{q}(K)$.

We shall say that a direction l in E^n sees a set $N \subset \operatorname{bd} K$ (or N is visible in a direction l) if for any point $y \in N$ the ray l_y with the apex y having the direction opposite to l does not intersect int K. A family $\mathcal{L}$ of directions sees K provided for any point $y \in \operatorname{bd} K$ there is an $l \in \mathcal{L}$ which sees y.

Denote by $w(K)$ the least number of directions in $E^n \setminus K$ which sees the whole body K. Obviously, $w(K) \leqq r(K)$. It is easy to see that $w(K) = 2$ if K is compact, and $w(K) = 1$ if K is unbounded.

Let $\overline{w}(K)$ denote the supremum of cardinalities of primitive families of directions which see K. As in the proof of Theorem 7, we obtain $\overline{w}(K) \geqq 1$ with $\overline{w}(K) = 1$ if and only if K is a half-space.

CONJECTURE 4. *If a convex body $K \subset E^n$ is compact, then $\overline{w}(K) \geqq n+1$.*

Note that $\overline{w}(K) \leqq 3$ for any convex figure $K \subset E^2$, and $\overline{w}(K) \leqq 2$ if K is unbounded. The value $\overline{w}(K)$ can be infinite in case $n \geqq 3$: for the convex body Q in Example 2, one has $\overline{w}(Q) = \infty$.

REFERENCES

[1] FEJES TÓTH, L., Illumination of convex discs, *Acta Math. Acad. Sci. Hungar.* **29** (1977), 355–360. *MR* **57** #4002

[2] HADWIGER, H., Ungelöste Probleme Nr. 38, *Elem. Math.* **15** (1960), 130–131.

[3] BOLTJANSKIĬ, V. G. and SOLTAN, P. S., *Combinatorial geometry of various classes of convex sets*, Shtiintsa, Kishinev, 1978 (in Russian). *MR* **80g**:52001 (See also: Combinatorial geometry and convexity classes, *Uspehi Mat. Nauk* **33** (1978), no. 1 (199), 3–42. *MR* **58** #7392.)

[4] BOLTJANSKIĬ, V. G., A problem about the illumination of the boundary of a convex set, *Izv. Moldavsk. Filiala Akad. Nauk SSSR*, No. 10 (1961), 79–86 (in Russian).

[5] GRÜNBAUM, B., *Convex polytopes*, Pure and Applied Mathematics, Vol. 16, Interscience Publishers, New York, 1967. *MR* **37** #2085

[6] KLEE, V., Extremal structure of convex sets, *Arch. Math.* **8** (1957), 234–240. *MR* **19**-1065

[7] EWALD, G., LARMAN, D. G. and ROGERS, C. A., The directions of the line segments and of the r-dimensional balls on the boundary of a convex body in Euclidean space, *Mathematika* **17** (1970), 1–20. *MR* **42** #5161

[8] DE WILDE, M., Some properties of the exposed points of finite dimensional convex sets, *J. Math. Anal. Appl.* **92** (1983), 257–264.

[9] VALENTINE, F. A., Visible shorelines, *Amer. Math. Monthly* **77** (1970), 146–152. *MR* **41** #2530

[10] BUCHMAN, E. and VALENTINE, F. A., A characterization of the parallelepiped in E^n, *Pacific J. Math.* **35** (1970), 53–57. *MR* **42** #5157

[11] BUCHMAN, E. and VALENTINE, F. A., External visibility, *Pacific J. Math.* **64** (1976), 333–340. *MR* **55** #11149

(Received July 23, 1990)

MATHEMATICAL INSTITUTE OF THE
MOLDAVIAN ACADEMY OF SCIENCES
STR. ACADEMIEI 5
KISHINEV 277 028
MOLDOVA

Studia Scientiarum Mathematicarum Hungarica

Volume 28

Akadémiai Kiadó, Budapest

1993

STUDIA SCIENTIARUM MATHEMATICARUM HUNGARICA

Volume 28

CONTENTS

ABBOTT, H. L. and ZHOU, B., On k-critical graphs with many edges and no short cycles 349
ANH, P. N., Rings with local units and descending chain condition 285
ARGYROS, I. K., A mesh-independence principle for nonlinear operator equations in Banach space and their discretizations 401
ARGYROS, I. K., Newton-like methods and nondiscrete mathematical induction 417
ASLAM, M. and ZAIDI, A. M., Matrix equation in radicals 447
BELL, H. E. and KLEIN, A. A., Two commutativity problems for rings 159
БЕРМАН, Д. Л., К теории экстремальных полиномиальных операторов 63
BIHARI, I., A generalization of the Riccati equation 79
BOGNÁR MÁTHÉ, K., Über Kugelsysteme unter Geräumigkeitsbedingungen .. 431
BOOK REVIEW 229
CHAJDA, I. and CZÉDLI, G., Mal'tsev functions on small algebras 339
CHUNG, P. V., Note on additive functions satisfying some congruence property. I 359
CHUNG, P. V., Note on additive functions satisfying some congruence property. II 427
CZÉDLI, G. and CHAJDA, I., Mal'tsev functions on small algebras 339
DEÁK, J., Extending a quasi-metric 105
FÉNYES, T., A note on a nonlinear operational differential equation system .. 465
FÉNYES, T., A note on the algebraic derivative and integral in a discrete operational calculus 457
FÉNYES, T., On an algebraic differential equation of Bernoulli type 115
FÉNYES, T., On an application of a binomial series expansion of discrete operators 231
FÉNYES, T., On the Fourier transform of the Bessel function with respect to the order 197
FÉNYES, T., On the Fourier transform of the modified Bessel function with respect to the order 189
GONCHIGDORZH, R., Generalized p.p. rings and rings of π-regular quotients .. 1
HAMEDANI, G. G., Characterizations of Cauchy, normal, and uniform distribution 243
HÉTHELYI, L., On the action of p'-automorphisms on p-groups having soft subgroups 317
JOÓ, I. and PALKO, M., On the directional derivatives 261
JOÓ, I. and SZABÓ, S., On the estimate of $(x_{\min} + x_{\max})/2$. II 321
JOOS, K., Nonuniform convergence rates in the central limit theorem for martingales 145
KENT, R. E., Dialectical logic: the process calculus 17

KHAN, L. A., Integration of vector-valued continuous functions and the Riesz representation theorem 71
KLEIN, A. A. and BELL, H. E., Two commutativity problems for rings 159
KOCA, K., Über eine Darstellung von Lösungen der komplexen Differentialgleichung $w_{\bar{z}}=b_{(F,G)}\overline{w}$ 387
KOMLÓS, J., REJTŐ, L. and TUSNÁDY, G., Learning with finite memory 173
KOVÁCS, K., On a conjecture of Kátai 237
KOVÁCS, K. I., On the convergence of the Fourier series of L-almost periodic functions 249
KROTOSZYŃSKI, S., Covering a disk with smaller disks 277
KY, N. X., On approximation by trigonometric polynomials in L_u^p-spaces 183
LAI, M.-J., Some sufficient conditions for convexity of multivariate Bernstein–Bézier polynomials and box spline surfaces 363
LASSAK, M. and VÁSÁRHELYI, É., Covering a plane convex body with negative homothetical copies 375
NGUEN, M. H. and SOLTAN, V. P., Lower bounds for the numbers of extremal and exposed diameters of a convex body 99
PALKO, M. and JOÓ, I., On the directional derivatives 261
REJTŐ, L., KOMLÓS, J. and TUSNÁDY, G., Learning with finite memory 173
RENDER, H., On the product of k- and l-spaces 453
SAKAI, R. and VÉRTESI, P., Hermite–Fejér interpolations of higher order. III 87
SAKAI, R. and VÉRTESI, P., Hermite–Fejér interpolations of higher order. IV 379
SARAN, J., Compositions of an integer and distributions of rank order statistics 267
SARIGÖL, M. A., A note on summability 395
SEBESTYÉN, Z., Restrictions of adjoint operators in Hilbert space 179
SOLTAN, V., External illumination according to L. Fejes Tóth 473
SOLTAN, V. P. and NGUEN, M. H., Lower bounds for the numbers of extremal and exposed diameters of a convex body 99
STEINFELD, O., Semigroups (rings) having a primitive regular (completely semi-simple) ideal 215
SZABÓ, S. and JOÓ, I., On the estimate of $(x_{\min}+x_{\max})/2$. II 321
TUSNÁDY, G., KOMLÓS, J. and REJTŐ, L., Learning with finite memory 173
VÁSÁRHELYI, É., Covering of a triangle by homothetic triangles 163
VÁSÁRHELYI, É. and LASSAK, M., Covering a plane convex body with negative homothetical copies 375
VERMES, I., Über die synthetische Behandlung der Krümmung und des Schmiegzykels der ebenen Kurven in der Bolyai–Lobatschefskyschen Geometrie 289
VÉRTESI, P. and SAKAI, R., Hermite–Fejér interpolations of higher order. III 87
VÉRTESI, P. and SAKAI, R., Hermite–Fejér interpolations of higher order. IV 379
VÉRTESI, P. and XU, Y., Truncated Hermite interpolation polynomials 205
WALENDZIAK, A., Join decompositions in lower continuous lattices 131
WINKLER, R., Polynomial approximation on locally compact abelian groups 135
XU, Y. and VÉRTESI, P., Truncated Hermite interpolation polynomials 205
YADAV, S. P., Saturation orders of some approximation processes in certain Banach spaces 299
ZAIDI, A. M. and ASLAM, M., Matrix equation in radicals 447
ZHOU, B. and ABBOT, H. L., On k-critical graphs with many edges and no short cycles 349

A New Mathematical Series

BOLYAI SOCIETY MATHEMATICAL STUDIES

The János Bolyai Mathematical Society has launched a new mathematical series called "BOLYAI SOCIETY MATHEMATICAL STUDIES" aimed to be a sort of continuation of the terminating old series **"Colloquia Mathematica Societatis János Bolyai"** published jointly with North-Holland. The scope of the volumes has been widened: they are not restricted any more only to conference proceedings, rather we aim to publish survey volumes or books; by all means, definitely more up-to-date and higher quality materials. Keeping this in mind, the first three books of the series are the following:

Volume 1: ***Combinatorics, Paul Erdős is Eighty, 1,*** published in July 1993

- 26 invited research/survey articles, list of publications of Paul Erdős (1272 items), 4 tables of photos, 527 pages

Volume 2: ***Combinatorics, Paul Erdős is Eighty, 2,*** to appear in Spring

- invited research/survey articles, biography of Paul Erdős

Volume 3: ***Extremal Problems for Finite Sets,*** to appear in Spring

- 22 invited research/survey articles

A ***limited time discount*** is offered for purchase orders received by ***March 31, 1994.***

Price table (US dollars)	Vol 1	Vol 2	Vol 1+Vol 2	Vol 3
(A) List price	100	100	175	100
(C) Limited time discount (purchase order must be received by March 31, 1994)	59	59	99	59

For shipping and handling add $5 or $8/copies of book for surface/air mail.

To receive an order form or detailed information please write to:

J. BOLYAI MATHEMATICAL SOCIETY,
1371 BUDAPEST, PF. 433, HUNGARY, H-1371
E-mail: H3341SZA@HUELLA.BITNET

Typeset by TypoTEX Ltd., Budapest
PRINTED IN HUNGARY
Akadémiai Kiadó és Nyomda Vállalat, Budapest

RECENTLY ACCEPTED PAPERS

SARAN, J. and RANI, S., Some distribution results on two-sample rank order statistics for unequal sample sizes

ALARCON, F., ANDERSON, D. D. and JAYARAM, C., Some results on abstract commutative ideal theory

FUCHONG, X., Order of best approximation by polynomials in H_q^p $(p > 0,\ q > 1)$

BOOTH, G. L and VELDSMAN, S., Special radicals of near-rings and Γ-near-rings

AKKOUCHI, M., Sur certaines algèbres associées à une mesure de Guelfand

ANGRISANI, M. and CLAVELLI, M., An alternative condition of fixed point of non-continuous mappings in metric spaces

CLAY, J. R. and YEH, Y.-N., On some geometry of Mersenne primes

CONTENTS

FÉNYES, T., On an application of a binomial series expansion of discrete operators 231
KOVÁCS, K., On a conjecture of Kátai 237
HAMEDANI, G. G., Characterizations of Cauchy, normal, and uniform distributions 243
KOVÁCS, K. I., On the convergence of the Fourier series of L-almost periodic functions 249
JOÓ, I. and PALKO, M., On the directional derivatives 261
SARAN, J., Compositions of an integer and distributions of rank order statistics 267
KROTOSZYŃSKI, S., Covering a disk with smaller disks 277
ANH, P. N., Rings with local units and descending chain condition 285
VERMES, I., Über die synthetische Behandlung der Krümmung und das Schmiegzykels der ebenen Kurven in der Bolyai–Lobatschefskyschen Geometrie . 289
YADAV, S. P., Saturation orders of some approximation processes in certain Banach spaces 299
HÉTHELYI, L., On the action of p'-automorphisms on p-groups having soft subgroups 317
JOÓ, I. and SZABÓ, S., On the estimate of $(x_{\min} + x_{\max})/2$. II 321
CHAJDA, I. and CZÉDLI, G., Mal'tsev functions on small algebras 339
ABBOT, H. L. and ZHOU, B., On k-critical graphs with many edges and no short cycles 349
CHUNG, P. V., Note on additive functions satisfying some congruence property. I 359
LAI, M.-J., Some sufficient conditions for convexity of multivariate Bernstein–Bézier polynomials and box spline surfaces 363
LASSAK, M. and VÁSÁRHELYI, É., Covering a plane convex body with negative homothetical copies 375
SAKAI, R. and VÉRTESI, P., Hermite–Fejér interpolations of higher order. IV 379
KOCA, K., Über eine Darstellung von Lösungen der komplexen Differentialgleichung $w_{\overline{z}} = b_{(F,G)}\overline{w}$ 387
SARIGÖL, M. A., A note on summability 395
ARGYROS, I. K., A mesh-independence principle for nonlinear operator equations in Banach space and their discretizations 401
ARGYROS, I. K., Newton-like methods and nondiscrete mathematical induction 417
CHUNG, P. V., Note on additive functions satisfying some congruence property. II 427
BOGNÁR MÁTHÉ, K., Über Kugelsysteme unter Geräumigkeitsbedingungen .. 431
ASLAM, M. and ZAIDI, A. M., Matrix equation in radicals 447
RENDER, H., On the product of k- and l-spaces 453
FÉNYES, T., A note on the algebraic derivative and integral in a discrete operational calculus 457
FÉNYES, T., A note on a nonlinear operational differential equation system .. 465
SOLTAN, V., External illumination according to L. Fejes Tóth 473

HU ISSN 0081-6906